Improving on the Standard Prospective Economic Evaluation Question

Marco A. Morales

The Wilf Family Department of Politics
New York University
19 W 4th St, 2nd floor
New York, NY 10012
marco.morales@nyu.edu

December 20, 2013

Abstract

Economic voting is perhaps one of the most studied relationships in political science. The last 30 years of research on economic voting based on survey data have employed a question that is ill-suited to study the relationship between the economy and vote choice in a manner that would be consistent with the theory. Focusing on *prospective sociotropic economic assessments* (*i.e.* evaluations about the future state of the national economy), the paper dissects the question that is typically used in survey-based economic voting analyses and singles out two specific causes of measurement error: the phrasing and the response scale of the commonly used question. It also proposes a corrected survey instrument - based on survey experiments conducted in the U.S. and Mexico - that consists of a battery of questions that condition on each candidate being in office, and utilizes response scales that disambiguate what respondents mean when they respond that the economy is the same as it is now.

Economic voting is perhaps one of the most studied relationships in political science. In short, economic voting is the common denomination given to the link between economic conditions and vote choice. Its appeal relies on the parsimony of its rationale: voters would keep an incumbent in office under whose stewardship the economy has performed well, or vote for a challenger if they think she could preside over a better-performing economy.

While empirical evidence has confirmed the existence of a strong link - at the aggregate level - between economic performance and vote choice over the last 50 years, the literature has not yet reached a consensus as to the precise mechanism that governs this relationship (Lewis-Beck & Stegmaier 2000, Anderson 2007, Linn, Nagler & Morales 2010). In particular, the link becomes less clear as we analyze the relationship between economic conditions - including their perceptions - and vote choice at the individual level, which typically relies on survey data of some sort.

This paper is about measurement error generated by a question that is ill-suited to study the relationship between the economy and vote choice in a manner that is consistent with the theory, and about the proposed means to minimize it from the question design stage.

In particular, this paper seeks to correct two main sources of measurement error embedded *by design* in the question typically used in surveys to obtain assessments about the state of the economy. The first one arises from using a *single* question that is unrelated to the candidates running in an election, when standard economic voting theory would require voters to compare *multiple* assessments about the state of the economy linked to each competing candidate. The second one results from using a response scale with a mid-category whose phrasing blurs the directional ("good" vs "bad") nature of respondents' assessments about the state of the economy when they consider the economy to remain unchanged from a previous state that is in itself directional.

I argue that it is paramount to address *both* sources of measurement error in order to utilize the best possible measures of assessments of economic conditions in surveys that are fit to conduct empirical tests of economic voting with the minimal amount of measurement error, at least from the design end. If both can be easily identified as sources of measurement error, and both can be corrected when designing the survey, it makes little sense to only correct one of them if it would be just as easy to correct them both simultaneously.

Recent literature in survey methodology has started to incorporate question design as part of a more comprehensive paradigm: the Total Survey Error approach (Weisberg 2005, Bautista 2012). An important part of this new comprehensive way of thinking about survey measures has question design as one important element (Krosnick & Presser 2010). A more careful thinking about the questions researchers ask and how they relate to the information they seek to obtain has started to advance the literature to the point of even correcting the way we measure traits that have been taken for granted by previous literature (Huber & Paris 2013).

Focusing specifically on *prospective sociotropic economic assessments* (*i.e.* evaluations about the future state of the national economy), the paper dissects the question that is typically used in survey-based economic voting analyses and singles out these two specific causes of measurement error. It also proposes a corrected survey instrument supported by evidence generated by experiments carried out in the U.S. and Mexico since 2008 that corrects this measurement error from the design stage.

1 Economic Voting: the standard measures

The study of economic voting surged after Kramer's (1971) foundational study linking aggregate economic conditions with congressional elections, later to be followed by Tufte (1975) looking at the link with presidential elections. Since then, a plethora of empirical studies have found a strong relationship between the economy and vote choice at the aggregate level (Goodman & Kramer 1975, Fair 1982, Hibbs 1987, Erikson 1989, Kiewiet & Rivers 1985, Lewis-Beck & Rice 1992).

Starting with Campbel et al. (1960), scholars have looked at the relationship between *perceptions* about the state of the economy and vote choice. But it was Fiorina (1981) who first presented an empirical test for economic voting based on key theoretical insights from Downs (1957), who had formalized the foundations for economic voting years earlier. From that point on, scholars have tried to determine empirically whether voters mark their ballots based on past (*retrospective*) or future (*prospective*) economic conditions. Evidence has been found that supports the view that voters rely on past economic conditions (Lanoue 1994, Alvarez & Nagler 1995, Krause 1997, Alvarez & Nagler 1998, Nickelsburg & Norpoth 2000, Alvarez, Nagler & Bowler 2000, Alvarez, Nagler & Willette 2000), but convincing evidence has also been presented of voters that primarily look at expected economic conditions to base their votes on (Lockerbie 1991, MacKuen, Erikson & Stimson 1992, Erikson et al. 2000, Lewis-Beck 1988).

Similarly, scholars have tried to determine if it is national (*sociotropic*) or personal (*pocketbook*) economic conditions that voters care more about when casting their votes. Just as in the previous dichotomy, evidence has been found supporting the idea that voters rely on their own economic conditions (Whitten & Palmer 1999), and similarly that national economic conditions dominate in voters' calculations (Kinder & Kiewiet 1981, Kiewiet 1983,

Markus 1988, Lockerbie 1991).

Conflicting results notwithstanding, these studies share theoretical underpinnings. First, they subscribe to the idea that voters differentiate between good and bad economic conditions and that this very difference between a good and a bad economy is the main driver behind their decision to select one candidate over the other(s). Second, they also subscribe to the idea that voters compare observed or expected economic conditions under the stewardship of each candidate. As will become clear shortly, the standard measures not only fail to unambiguously capture these traits, but add a degree of confounded measurement error that derives from their phrasing and design.

1.1 Two sources of measurement error in the standard measure

Studies of economic voting based on individual-level survey data utilize instruments that embed measurement error crafted in them at the design stage. The first source of measurement error arises from asking a general question that omits any reference to the candidates competing in the election. Seminal theories of economic voting (Downs 1957, Fiorina 1981) pose that voters generate utility assessments linked to each candidate that derive from their (observed or expected) economic performance while in office. Voters would compare these assessments and select the candidate associated to the highest expected utility. Nevertheless, the last 30 years of individual-level analyses have relied on survey data that uses a *single* measurement about the state of the economy dissociated from any candidate, rendering any comparison impossible. This can be easily corrected simply by asking as many questions as there are candidates competing in the election, and conditioning each question on each candidate being in office. This allows researchers to model explicitly the comparisons between the economy associated to each candidate that is posited by standard economic voting

theory.

The second source of measurement error arises from utilizing a response scale with a logically ambiguous mid-category - "the economy will be the same as it is today" - which ignores that respondents must have a prior assessment about the state of the economy today that would lead them to choose this category. More precisely, respondents who think that the economy is "bad" today and think it will remain as such, would be lumped together with respondents who think that the economy today is "good" and will remain so, when the first one should logically belong to the negative side of the scale, and the second one to the positive one. Adding to the gravity of the argument, and as will be shown later in the paper, the mid-category in economic assessment questions typically does not have a negligible share of respondents. This should be of concern to researchers as economic voting theory predicts different consequences from each one of these interpretations. Hence, information is lost due to an ambiguous mid-category that obscures these differences and generates the known problems for econometric estimation derived from this engrained measurement error. It can be easily corrected by replacing the confusing mid-category "same" with two explicit interpretations of this response that ground it on respondents current assessments: "as good as it is today" or "as bad as it is today" which would minimize this measurement error. This allows researchers to fully exploit information that is relevant for testing economic voting theory.

From the discussion above, it should be clear that the commonly-used question is *not* the best question to be used to test economic voting, and that it generates specific kinds of measurement error *by design*.

1.2 Measurement error in the economic voting literature

The literature has not been silent regarding measurement error in surveys that contain questions that are typically used in economic voting studies. For example, Wilcox & Wlezien (1993) show that assessments about the state of the economy can be contaminated by other political attitude questions placed in the same survey. Similarly, Palmer & Duch (2001) show that assessments about the state of the economy are influenced by the placement of the question in the survey, or by cues in the survey itself.

On a related perspective, researchers have also been concerned with endogeneity in assessments about the state of the economy derived from party preference (Wlezien, Franklin & Twiggs 1997, Evans & Pickup 2010, Gerber & Huber 2010, Tilley & Hobolt 2011, Chzhen, Evans & Pickup forthcoming), likelihood of a party winning the election (Ladner & Wlezien 2007), whether a party won the election or not (Anderson, Mendes & Tverdova 2004, Glasgow & Weber 2005), prior opinion of the incumbent (Evans & Andersen 2006), or government approval (van der Eijk et al. 2007, Pickup & Evans 2013) to name just a few.

It is just recently that measurement errors derived from the phrasing and/or scales used in the questions typically used in surveys to inquire about the state of the economy, which are the workhorse of economic voting analyses that rely on survey data, have begun to be explored. In a recent paper that critiques the naïve prospective economic evaluation question in a pre-election context, Michelitch et al. (2012) argue, among other things, that it is nearly impossible to know *a priori* how to interpret the information obtained with this question, and suggest a series of possible heuristics that respondents might employ to answer it.

When measuring econonomic assessments in general, it would be desirable to have measurements that - as the theory dictates - differentiate between the economic situation

that individuals think will prevail if each of the candidates were in office. But also that they clearly determine a point of reference that can unambiguously establish change in these conditions relative to a given point in time, and that are measured using a scale that captures the latent variable of interest with the least possible amount of measurement error. Those would be minimal requirements regarding question design to mitigate the measurement error described above.

Survey experiments were conducted in the U.S. and Mexico to empirically evaluate solutions to these problems in the commonly used questions. The evidence presented in this paper, leads me to conclude that one plausible way to capture the information we would need to more accurately test economic voting, and reduce measurement error is by using the questions proposed here. These instruments are a battery of questions, where each is conditioned on a particular candidate being in office. But also, utilize a branched response scale that allows respondents to disambiguate the meaning of their answers and better ground their responses on a specific time frame. This would bring us much closer to what we would ideally require as valid measures of economic evaluations to be used in economic voting studies.

As means to focus the discussion, and explain the arguments in detail, this paper will focus exclusively on *sociotropic prospective economic evaluations*. This does not imply that the arguments of this paper cannot be extended to *retrospective* evaluations, and similarly to *pocketbook* ones. On the contrary, minor adjustments in the *phrasing* of the proposed questions would suffice to make them suitable to examine these variants of economic voting. The key points for doing so, are addressing the two sources of measurement error that are addressed here.

2 Correcting measurement error in question phrasing

Let us begin by steps, addressing first measurement error caused by the phrasing of the question. As mentioned earlier, seminal theories of economic voting (Downs 1957, Fiorina 1981) posit essentially that voters engage in an exercise of comparison between all candidates competing in an election, and choose to vote for the one that produces the highest expected utility. The state of the economy is also posed as the main area of assessment for each candidate. Hence, voters would imagine a state of the economy that would prevail if a specific candidate were in office, and compare these expected assessments.

It would be expected that surveys utilized in economic voting studies should gather this information from respondents. Unfortunately, economic voting studies do not utilize multiple assessments about the state of the economy that correspond to each candidate being in office. Instead, they utilize information from a generic question whose phrasing is typically close to what the American National Election Studies (ANES) asks:

> What about the next 12 months? Do you expect the economy, in the country as a whole, to get BETTER, stay ABOUT THE SAME, or get WORSE?

Note that, contrary to what the theory requires, the question asks for a *single* assessment about the state of the economy that is *not* explicitly conditional on any candidate being in office in the future. This last feature is a requirement of the theory to be able to compare the expected performance of the economy under each candidate. Furthermore, it gets in the way of operationalizing the comparison posed by the theory, which an adequate econometric specification should mimic (Alvarez & Nagler 1995, 1998).

More worrisome for researchers engaged in economic voting should be to know what information do respondents provide when asked the naïve question. In the previous paper

of this dissertation, I show that prior to the election, respondents answer the naïve question with an assessment about the future state of the economy under their preferred party with some qualifications: the incumbent party if they prefer that party or the challenger when they prefer that party *and* think it will win the election. After the election has taken place, they converge to providing assessments about the future state of the economy under the party that won the election. Fundamentally, since the naïve measure combines different pieces of information into a very noisy measure, it fails to obtain the information it should from respondents.

It should be evident by now that the question that is typically used in economic voting studies must be corrected. One appropriate way to do it is by replacing it with a battery of questions, each of which should explicitly condition on a particular candidate being in office. This should be a first step to correct measurement error embedded in the design of the survey question.

Even when the theory assumes that voters produce assessments about the (future) state of the economy under each candidate running in the election, a first concern for researchers should be to verify that survey respondents are capable of generating these assessments. If that is the case, we must then subject these responses to examination and assess whether they conform to what economic voting theory poses.

For this purpose, a survey experiment was carried in the 2008 and 2012 pre-election editions of the American National Election Studies (ANES). It consisted of adding two new questions on economic assessments - each one conditional on one of the candidates winning the election - in addition to the naïve question. (See Appendix A for details.)[1] The questions were phrased as follows, substituting [CANDIDATE] for either "BARACK OBAMA" and "JOHN McCAIN" in 2008, or "BARACK OBAMA" and "MITT ROMNEY" in 2012:

> What about the next 12 months? If [CANDIDATE] wins the election, do you expect the economy, in the country as a whole, to get BETTER, stay ABOUT THE SAME, or get WORSE?

In essence, what the inclusion of the phrase "If [CANDIDATE] wins the election" does is *priming* respondents so that their answers retrieve specific information from memory on that particular candidate and relate it to the expected performance of the nation's economy. Instead of conceiving priming as a problem to be corrected, it is used here as means to prompt survey respondents to think about the question in a particular manner as a strategy to obtain specific information from them (Hopkins & King 2010). In this particular case, priming is used to obtain information that more closely complies with what economic voting theories would require.

As a result of this new phrasing, the conditional questions have two main improvements over the naïve one. First, they anchor the assessment of the future state of the economy on a particular candidate being in office. And, second, they generate two measurements - one for each candidate competing in the election - enabling researchers to explicitly model their comparison.

Table 1 summarizes the distribution of responses to all three questions for the 2008 ANES experiment. Note that the distributions are quite different across the two conditions. If we were to look only at the responses to the naïve question, we would naturally believe that Americans foresee the economy over the next year to be the same as it is now (38%) or improving (34%). Nevertheless, if we look at the conditional questions we see a different story. Americans, in general, thought prior to the election that the economy would perform better under Obama (47%) or remain the same (38%). But also, that under McCain it would remain similar (50%) or worsen (32%). These are two very different pictures of the future economy derived exclusively from conditioning - as the theory would have us do - on

a particular candidate being in office.

[Table 1 about here]

Similarly, Table 2 presents response distributions for all three questions on the 2012 ANES.[2] Just as in the previous case, the distributions vary notoriously across all three questions. If we were to look only at responses to the naïve question, we would conclude that voters foresee an economy that would improve (46%) or stay about the same (40%). But the conditional questions show that Americans have a bit more polarized forecasts of the economy if each candidate were elected President. Under a second Obama Administration, they mainly foresee an improving economy (48%), although they foresee an either worsening (39%) or improving (36%) economy if Mitt Romney were elected president.

[Table 2 about here]

These are troublesome findings, as any analysis that uses the naïve questions would erroneously reach conclusions on prospective economic voting based on measurements that suggests that people think that the economy would move in one particular direction. But we know now, thanks to the conditional questions, that people can and do generate prospective economic assessments conditional on each candidate being in office that are quite different from each other, and from the naïve one. All of this, derived from measurement error caused by the question that is commonly used. See Figure 1 for a visual representation of these differences.

[Figure 1 about here]

The same evidence strongly suggests that people may be combining - in some manner - their conditional assessments on both candidates when answering the naïve question as a

result of the phrasing of the question alone. If this is the case, researchers have been using a very noisy measure to study (prospective) economic voting. This fact could help explain why there is so much inconsistency in results across elections and even within a same election depending on the time in which the questions are answered.

Thus far, the evidence presented here has shown that respondents are able to generate the conditional assessments about the state of the economy that economic voting theory assumes they do. But do these assessments also conform to the relationship with vote choice that the theory also proposes? We do know, in fact, that one can find very strong prospective economic voting effects for the 2008 US election if one uses the conditional questions, and - not surprisingly - that there are no prospective economic voting effects if the naïve question is used (Michelitch et al. 2012). The conditional measures perform better because they are able to parse out *explicitly* an assessment for each one of the candidates competing in an election, and produce results that more closely resemble those that would be expected by the standard theories of voting.

Armed with this knowledge, we can proceed to address the next source of measurement error generated by what I argue is a faulty response scale. The rest of the paper will present the results of survey experiments conducted in Mexico for this purpose.

3 Correcting measurement error in the response scale

The next step, then, is to correct the response scale that is typically associated to economic assessment questions. Let us begin conceptually. We could think of the state of the economy as a bipolar (better/worse) construct *without* a natural midpoint category. It would be logical to do so since a midpoint in the response scale labeled "same" must be a function of

the state of the economy in a previous period that serves as a baseline for the comparison. And logically, the state of the economy in this previous point in time must also be bipolar.

In other words, when someone claims that the economy will be the same as it is now, she must first know how the economy is today. And that assessment can only be positive or negative. So saying that the economy will remain the same as it is today translates into saying that the economy will be good or bad, depending on the assessment about the economy today. This matters because even if choosing the mid-category does not imply a *change* in the state of the economy (better/worse), it most certainly implies a *direction* (positive/negative) in the assessment of the economy today and in the future that is relevant for economic voting purposes. In other words, the mid-category is not neutral. It does convey some information on the expected *direction* of the state of the economy.

For this reason, it can only be natural to decompose any mid-category into the natural two categories (good/bad). That renders an even-numbered response scale that allows researchers to extract all the information from respondents that is relevant for economic voting (Kroh 2007). In this case, having response categories that correspond to the natural ordering of the latent responses may assist to disambiguate an unclear phrasing of the question (Schwarz 1990, Schwarz, Grayson & Knäuper 1998).

It will be obvious by now that this mid-category simply lumps together positive and negative assessments. Despite this fact, response scales for economic assessment questions typically include a mid-category, and no disambiguation follow-up question. This alone would pose a problem for researchers, as what we are trying to measure are precisely these differences between good and bad in the assessments.

Phrasing the mid-category in an ambiguous manner could be inadvertently facilitating *satisficing* behavior on the part of respondents, instead of a more desireable *optimizing*

(Krosnick 1991). This would happen because respondents who don't strictly think that the economy would improve (worsen), but that would remain as good (bad) as it is now, face the lack of an adequate response category in the scale. So, when asked the question, they would have to either opt for the "same" category even when it does not clearly distinguish between "just as good" and "just as bad", or refuse to respond the question. But the more problematic behavior would be that, seeking to avoid the additional cognitive effort required to disambiguate the meaning of the question given the response scale, they would simply choose the mid-category to provide an answer to the question, even when it does not reflect their true assessment. If they were provided with other alternatives in the scale that better captured their assessment, they would most likely choose it.

We could take two paths to disambiguate the mid-category. On the one hand, we could not offer the "same" category upfront in the response scale. Eliminating ambiguous items in the response scale, increases its reliability and interpretability (Coombs & Coombs 1976), and allows a researcher to better capture the information that is sought. On the other hand, we could prompt respondents to refine their answer if and when they choose a "same" response, so that we can clearly know whether they mean "just as good" or "just as bad" as it is now. Question branching of this sort provides respondents with additional time to retrieve information from memory, and to better cope with the additionl cognitive demands of the question as phrased. Similarly, when faced with ambiguous questions, further pressing the interviewee to provide a more truthful answer can increase the reliability of their responses (Krosnick & Presser 2010). The experiments presented on this section exploit the advantages of this second path. In particular, they seek to explore the appropriateness of decomposing the mid-category, and finding the best scale to do so.

A survey experiment was conducted in Mexico embedded in a survey conducted by the polling firm DEFOE, with a national probability sample (n=800) during September of

2008. A split-sample design randomly assigned each version of the question and its scale to a subsample of survey respondents, and can be described as follows (see Appendix B for details):

a) a **branching** condition where the response scale included five labeled categories: "Much better", "Better", "About the Same" "Worse", and "Much worse". If respondents choose the "About the Same" category, they would be prompted with a follow-up question to further qualify whether they meant "just as good" or "just as bad".

b) a **repeat question** condition where the same five response categories are offered. If a respondent chose the "About the Same" category, the interviewer would repeat the same original question a second time. If the respondent insisted in answering "About the Same", the interviewer would repeat the same original question a third time. Only if the respondent insisted in responding "About the Same" on all three occasions, would this answer be coded as such. The rationale of behind this tactic is to force the respondent to think about the question and avoid using the "About the Same" category unless she actually meant it.

The design of both conditions is exploited to tease out the information that might be confounded in these answers. Given that they share the exact same phrasing and initial response categories, the observed variation in response distributions would be solely due to manipulations in how responses that fall in the "About the Same" category are treated.

The **branching** condition requires respondents to clarify their interpretation of "About the Same" response category. Figure 2 shows that 41% of respondents answered "About the Same". After asking them to clarify their response, they favored "just as bad" in a 3:1 ratio over "just as good". This is a non negligible difference and suggests that our usual phrasing

masks and mixes information that would be useful for researchers seeking to parse out the effects of prospective economic assessments on vote choice.

[Figure 2 about here]

While the **branching** condition decomposes the "same" category into two additional categories, the **repeat question** condition seeks to reassign answers that do not strictly fall in the "same" category, while still maintaining this alternative as an option, iff offered by the respondent and sustained after three consecutive attempts by the interviewer. This is an original artifice designed to press respondents to think thoroughly about the answers they provide to this ambiguous question and single out respondents whose true attitude is that the economy will remain the same as it at the time the question is asked. It is inspired in strategies employed by survey methodologists and social psychologists who have shown in other contexts that pressing respondents to think further about their answers can increase the reliability of vote intention reports to surveys when pressed to think about and report the candidate towards whom they leaned (Visser et al. 2000), by prompting respondents for an answer whenever a blank response was generated on internet surveys (Fricker et al. 2005), or be instrumental to separate *real* "don't know" answers from *non-attitudes* by forcing respondents to identify whether they have information that is germane to the question they are asked or not (Norman 1973, Glucksberg & McCloskey 1981).[3]

Aligned with these expectations, figure 3 shows that when prompted for a more thoughtful answer, respondents shift the response distribution in favor of the "worse" category, just as they did in the previous condition. But note also that after forcing respondents to think harder about the question, still 26% of respondents chose to maintain an "About the same" answer, which is also not a negligible amount.

[Figure 3 about here]

Both conditions are not strictly comparable, as one renders a six-category response distribution and the other a five-category. Nevertheless, if the mid categories in the **branching** condition are bundled together again into their original "About the Same" response, we could have comparable distributions. Figure 4 compares them both. Note that the overlapping 95% confidence intervals suggest that both distributions are quite similar, thus suggesting that both would be originated by the same distribution.

[Figure 4 about here]

Readers may be wary that the response categories included in the previous experiments are "vague quatifiers" and thus lead to even more inaccurate responses (Tourangeau, Rips & Rasinski 2000). That is, the categories could be too "wide" to capture more precisely respondents' attitudes, so a more straightforward response scale could be appropriate. To account for this possibility, an additional experiment was conducted using a 5-point numeric scale where 1 was labeled "much worse" and 5 "much better".

[Figure 5 about here]

As can be seen in Figure 5, this particular response scale does not resemble the distribution in the previous response scales as it clusters in the positive assessments end of the scale. This finding is not unexpected, as respondents are known to cluster responses away from zero in numeric scales (Schwarz 1990, Schwarz et al. 1991, Schwarz 1995). Even if a numeric scale could be less "vague", it introduces a different type of measurement error absent from the previously tested scales.

Various things can be learned from these experiments. First, when respondents choose to provide a "same" answer, they are able to further clarify the meaning of this question and

determine whether they meant "as good" or "as bad" as it is today. That is, respondents do not seem to have a problem providing a "same" answer even when it wouldn't be logical to do so. But they are also capable and willing to clarify the meaning of this answer in a most logical manner. So the seemingly ilogical behavior on the part of respondents is not so, once a clarification of the answer is made.

Second, even after forcing respondents to think harder about the question and their possible answers, they seem inclined to answer "same" to assesments about the (future) state of the economy. This could be due to the fact that respondents who thought hard about their answer do not intend to signal change in the state of the economy and thus choose to respond "same".

Finally, we also learn that a labeled scale is more efficient than a numeric scale to capture the information we seek to obtain regarding assessments about the (future) state of the economy.

How do we translate this new information into a more suitable scale to better measure prospective economic evaluations? We may want respondents to offer a "same" response only when they have one. The natural way to make sure of this could be by provinding it as a category in the response scale, as many people would want to see it there. But, when respondents decide to provide this answer, we may also want its meaning clarified and know whether it means "just as good" or "just as bad".

Hence, a better response scale could be branched, with a 5-category initial scale that includes "Much better", "Better", "About the Same", "Worse", and "Much Worse", followed by a 2-category scale composed of "As good as it is today" or "As bad as it is today" to be asked when respondents choose the "About the Same" category. In other words, if respondents do know what they mean by their answer, we just need to ask them. And the

best way to do so is using a response scale where "same" can be clarified as to its meaning: *just as good* or *just as bad*.

4 Conditional Questions with Unambiguous Response Scales: an Empirical Test

Having explored various alternatives for the response scale, the next natural step is to test how conditional questions coupled with the proposed scale fare among survey respondents. This experiment would show, primarily, whether it poses too demanding a cognitive task for them.

For this purpose, an experiment was conducted in Mexico during June 2012, just days before Election Day, embedded in a survey conducted by the polling firm DEFOE, with a national probability sample (n=1,200). The survey contained two experimental conditions. The first one, assigns a battery of three prospective economic assessment questions conditional on each one of the three main presidental candidates running in 2012. The second one, to serve as a control group, assigns the naïve question to respondents (see Appendix C for details). Both share the branched response scale that seeks to clarify the mid-category.

Table 3 shows the response distribution for both experimental conditions. Note that the response distribution for the naïve question is quite different from that for the conditional questions. Also, there are interesting differences within responses to the conditional questions. The one corresponding to PAN's candidate, Josefina Vázquez Mota (JVM), is unimodal over the negative end of the scale. In contrast, the other two questions conditional on PRI's Enrique Peña Nieto (EPN), and PRD's Andrés Manuel López Obrador (AMLO), are clearly bimodal over the better and worse categories.[4]

[Table 3 about here]

To restate the differences, the distribution of responses to the naïve question would tell us that prior to the election, most respondents had a very positive outlook for the Mexican economy. This contrasts starkly with the response distributions for the conditional questions where, conditional on having Vázquez Mota in office generates a bad economic outlook, but an either good or bad outlook conditional on Peña Nieto or López Obrador being in office. For added clarity, figure 6 provides a graphical representation of response distributions for both experimental conditions.

[Figure 6 about here]

In addition, these experiments also show that the measurement scale, as designed, can disambiguate the meaning of the "same" category. In both the conditional and naïve questions, the midcategories show a notable degree of variation. In particular, the distribution patterns in the conditional questions' midcategories differ from one another, suggesting a pattern that is not a mere random decomposition of the "same" category, but an intentional response that differentiates between categories.

Table 4 groups the proposed conditional economic assessment questions according to the vote intention of respondents for each one of the three main presidential candidates competing in the 2012 Mexican election to get a better sense of its relationship to vote choice.

[Table 4 about here]

Note first that there is a strong correspondence - just as the theory would suggest - between economic assessments about the future state of the economy under a given candidate,

and the likelihood of voting for that particular candidate. As shown by the percentages in bold, consistently over 60% of those who think the economy will improve ("much better" or "better") under a particular candidate would vote for him. Note also that consistently over 40% of voters who think the economy will decay ("worse" or "much worse") under a particular candidate tend to favor candidates other than that one. This strongly conforms to the central tenets of economic voting theory and should enhance confidence in the conditional aspect of the measures in multiparty settings as well.

But more interesting for the purposes of this paper is examining how the disambiguated mid-category responses perform relative to vote choice. That is, how does vote choice allocates among the totality of respondents who think that the economy will remain just as good (or just as bad) under a given candidate, and also relative to the ambiguous mid-category. Table 5 provides a much clearer picture of these dynamics, as it compares the distribution of intended vote choice among respondents whose assessments fall in the mid-category presenting them as disambiguated (as good/as bad) in the top half, and ambiguous (same) in the bottom half. Each row shows the percentage of votes that would go to each of the three candidates provided that the economy would remain just as good (or as bad) under each one of the competing candidates. Each row adds up to the totality of respondents that provided the row answer.

[Table 5 about here]

One could reach very different conclusions by looking only at the top or at the bottom parts of the table.[5] For example, the ambiguous measure would suggest that a plurality of respondents who think the economy will be the *same as it is today* under EPN would vote for AMLO. Yet, if we look at the disambiguated measure we see that this is *only* true for those who think the economy will be *as bad as it is today*, while those who think that the

economy will be *as good as it is today* will tend to vote primarily for JVM.

This could seem odd at first - respondents who think the economy will remain as good under EPN will *not* vote for him, and those who think it will remain as bad will also *not* primarily vote for him.[6] But we must remember that what we are looking at on table 5 is not how does vote fare among voters who think the economy would *change* (improve/worsen) from what it is today under a given candidate; those results are shown with clarity on Table 4. What we are looking at are vote distribution patterns when voters think that the economy will not change, but maintain the same *direction* (good/bad) it holds today under a given candidate. That said, it does makes sense to retain the incumbent party when the economy will be as good as it is under the current incumbent, while it also makes sense to choose a different challenger if the economy will be as bad under EPN, one of the challengers. Hence, these results are also logical if the economy remains the same.

Interesting patterns also arise in the cases of assessments conditional on candidates from the other two parties. If the economy will not change under the candidate of the incumbent party, voters will favor one of the challengers. But again, the devil is in the detail, as revealed by the disambiguated mid-category. Those who think the economy will remain as good under JVM *as good as it is today* will not primarily favor her. They would vote primarily for EPN. But among those who think the economy will remain *as bad as it is today*, they will split nearly by halves favoring EPN and AMLO. This last nuance is lost in the ambiguous category that suggests that voters who think the economy will remain *the same* will overwhelmingly prefer EPN.

Perhaps the least interesting case relates to assessments about the state of the economy under AMLO. Whether voters think the economy will be *as good*, *as bad*, or *the same* as it is today, they will overwhelmingly vote for EPN, or JVM in second place. Yet, interestingly enough, AMLO is consistently their least preferred alternative. And this may be due

to the fact that his party has never held stewardship of the country.

In essence, what having a disambiguated mid-category does is enhance our power to correctly classify between 1/4 and 1/3 of respondents who think the economy will not change, but maintain the same *direction* under a given candidate.

Given these results, we could expect that the disambiguated variable assists in better classify cases when estimating econometric models. Two simple vote choice models - otherwise identical but one using the disambiguated and the other the ambiguous measure - were estimated (see Appendix D for details) to assess a potential improvement in classification power from this measure. Table 6 presents the change in performance across models, and includes Percentage Correctly Predicted (PCP) and Herron's (1999) Expected Percentage Correctly Predicted (ePCP) statistics for both models.

[Table 6 about here]

In general terms, the model with the disambiguated prospective economic assessment variable seems to do a marginally better job of correctly classifying (78.98% vs 78.59%) vote choice. But if we look closer at each possible vote choice alternative j, a clearer pattern emerges. The model with the ambiguous measure does a slightly better job of classifying voting for EPN (88.73% vs 88.27%). But the model with the disambiguated measure improves a bit more notoriously classification for JVM (66.03% vs 64.09%) and AMLO votes (77.02% vs 76.05%).

In terms of ePCP, the model with the ambiguous measure performs negligibly better (62.66% vs 66.65%), but this is probably due to the marginal changes in probabilities of selecting each alternative j that results from utilizing the disambiguated measure in the estimated model. So perhaps looking at where these changes in probability happen might

be illustrative.

Figure 7 presents the difference in probabilities of selecting party j between the model that utilizes the disambiguated measure and the ambiguous one, specifically for respondents who think the economy would remain "just as good" or "just as bad" under candidate j. That is, distributions over the negative end of the scale denote that the model with the disambiguated variables assigns a *lower* probability of selecting alternative j under the specific conditional assessment than the model with the ambiguous variable. The opposite - a *higher* probability - is true of the positive end of the scale. Note also that the figure distinguishes between the probabilities assigned for voters who actually selected alternative j and those who did not.

[Figure 7 about here]

As would be expected, utilizing the disambiguated measure always increases the probability of selecting an alternative when respondents think the economy under party j would be "as good" as it is today, and reduce it when respondents think the economy will be "as bad" as it is today.

More interestingly, there is a notorious number of cases where there is a reduction in the probability of selecting an alternative that was *not* actually chosen by a respondent where she thought that the economy would be "just as bad" under party j. This was a bit les prevalent among voters who actually selected the alternative.

Among those who thought the economy would remain "just as good" under party j, there was an increase in the probability of selecting that alternative among those who actually selected that alternative in the model that uses the disambiguated question. Yet, the more prevalent case was an increase in probability of selecting alternative j among those

who did not vote for that alternative.

In sum, the model that utilizes the disambiguated measures, relative to the model that uses de ambiguous one, has two general effects. Most changes in probability seem to happen among respondents who did not vote for a given alternative, and reductions in the probability of selecting alternative j seem to be especially prevalent among respondents who think the economy will remain "as bad" under party j, and didn't vote for that party. Both could be helpful dynamics to improve - even marginally - estimation accuracy.

This becomes especially important in multiparty settings where, contrary to the bi-partisan case, a voter has multiple "fallback" candidates if she decides not to favor the candidate of the incumbent party. That is, if a voter thinks that the economy will perform well under candidate A, she might choose candidate A provided that she doesn't think that the economy will perform better under other candidates. But if she thinks that the economy will perform badly under this same candidate, she could either choose candidates B or C depending on how she expects the economy to perform under their tenure. What these results underscore is that there are interesting dynamics related to vote choice that would be masked if the question were just simply assumed to have a neutral midpoint in the scale.

One thing seems to be clear from these results that adds a degree of complexity to economic voting theory: *change* engenders different dynamics than *direction* among voters. When voters think that the economy will *change* for good under a particular candidate, they seem to favor that candidate, and if the change is for bad they will not favor the candidate. In contrast, when there is no change involved but preservation of the *direction* of the economy under a candidate (irrespective of whether it is good or bad), voters will tend to migrate to other candidates.

But if this intuition in correct - and the data suggests that it is - then it makes

little sense to include it as a single variable in models. It would make more theoretical and empirical sense to exploit information in this variable more fully by singling out each response category.

5 The proposed new measures

The evidence produced by the previous experiments provides guidance as to what an appropriate combination of a question and its measurement scale to assess the (future) state of the economy should look like. The proposed questions to be used in economic voting analyses based on survey data should comply with three basic requisites:

a) must be conditional on *every* candidate winning the election, therefore producing as many questions as candidates compete in the election,

b) must include a midpoint "same" response category, and

c) must include a follow-up question to disambiguate the meaning of a "same" response when provided by the respondent

Naturally, there is a difference in the information to be collected from respondents depending on whether the identitiy of the winner of the election is known. So a full proposal for new measures should have a pre-election and a post-election version. To better illustrate these differences, I exemplify the proposed questions as if they had been asked before an after an election where an incumbent president was running for reelection.[7] In the case of pre-electoral surveys, the question should be phrased as follows:

> Thinking about the economy over the next presidential term, if [{incumbent} were reelected/{challenger} were elected] as President, would you say the nation's

> economy will be better off, about the same, or worse off than it was during {incumbent's} current presidential term?

In consonance with what has been argued earlier, when a "same" response is offered, the question should be followed by a question to clarify the meaning of same, and be phrased as follows:

> You answered the nation's economy will be about the same. Does "About the same" mean as good as it was during {incumbent's} previous term, or does "About the same" mean as bad as it was during {incumbent's} previous term?

In the post-electoral surveys, once the winner of the election becomes known, the phrasing of the question should be modified accordingly to maintain a proper conditioning, to render questions phrased as follows:

> Thinking about the economy over the next presidential term, [now that {incumbent} has been reelected as President / now that {challenger} has been elected as President /if {incumbent/challenger} had been elected President], would you say the nation's economy will be better off, about the same, or worse off than it was during {incumbent's} current presidential term?

This phrasing guarantees that the questions are consistent with the theory, with the time in which they are asked, and utilize a response scale that minimizes the measurement error derived from question design.

6 Discussion

The results presented in this paper have a series of implications that are important to discuss in further detail. First, they show that respondents can provide assessments about the state

of the economy conditional on a given candidate being in office. When asked to, they are capable of retrieving information that is relevant to such assessment and provide it to an interviewer.

This could have potentially been an important obstacle for economic voting theory, since individuals that cannot generate assessments about the state of the economy if a certain candidate were in office, are also logically unable to compare them. If proven true, this would have severly undermined the central tenet that underlies theories and models of economic voting that are commonly used as foundations on the empirical literature.

But this also raises an additional number of interesting questions. It may be the case that individuals think that each candidate would be able to influence the economy in specific manners, so that it would perform differently under each candidate. But it could be equally possible that respondents recognize a natural growth rate for the economy that is independent of who is in office, and that could serve as a point of reference for changes due to economic management. This could be, of course, a matter to be settled empirically, but with potential effects in the way we conceive voters and their relationship to economic performance.

Secondly, we learned that respondents provide ambiguous economic assessments when the response scale is not sufficiently clear in what it seeks to capture. Respondents usually adapt to the response alternatives provided by the interviewer thinking that the researcher either knows best, or simply that she wants respondents to conform to that specific scale. So it is important to provide respondents with unambiguous scales that capture the information that we seek to collect.

This is an important finding, since the "same" category combines information about an improving and a worsening economy in the same response category. Economic voting

relies on being able to distinguish these two directions, since we think that each direction in the assessments would have differentiated effects for each candidate. In essence, we could be having varying degrees of bias in our estimates derived from this confounded information.

Fortunately, we also learned that it is possible to collect cleaner information from respondents if we ask them to clarify their answers with follow-up questions. That is, by asking them to state whether they mean that the economy is "as good" or "as bad as it is today" when they say the economy will remain the "same", we retreive directional assessments that are better suited to assess the effects of economic evaluations on vote choice.

Thirdly, the combination of a better phrased question, that asks respondents to condition their economic assessments on particular candidates being in office, with a response scale that can disambiguate the meaning of the mid-category ("same") provide researchers with responses that better allign measurements with the theory, and reduce measurement error in the responses.

In essence, a better measure for assessments about the state of the economy that is germane to economic voting theory should address two sources of measurement error: one derived from the phrasing of the question to make it conditional on a particular candidate being in office, and another derived from the response scale that the question utilizes so that all available information is extracted from survey respondents.

The evidence presented here corresponds to only one of four possible measurements that are typically advocated in the economic voting literature that uses survey instruments, namely prospective sociotropic evaluations. In principle, there is no good reason to justify not applying what we learned here for *sociotropic prospective* economic assessments to the remaining variants of questions that have been utilized to empirically test economic voting

theories: prospective pocketbook, retrospective sociotropic and retrospective pocketbook evaluations. They all share a theoretical framework that requires comparisons of assessmets *conditional* on each possible candidate being in office. Similarly, all these questions require being asked alongside a response scale that is clear an unambiguous. All of this with the purpose of reducing measurement error that is derived from an inadequate design of the questions used to test economic voting theories.

A question that may naturally arise on these proposed new measures is whether they suffer from some sort of endogeneity. As I pointed out earlier, endogeneity in economic assessments has been an important concern that has gained traction in the literature (Wlezien, Franklin & Twiggs 1997, Anderson, Mendes & Tverdova 2004, Glasgow & Weber 2005, Evans & Andersen 2006, Ladner & Wlezien 2007, van der Eijk et al. 2007, Evans & Pickup 2010, Gerber & Huber 2010, Pickup & Evans 2013, Chzhen, Evans & Pickup forthcoming). While the arguments that have been advanced are theoretically appealing, it is equally important to note that all the evidence that has been marshaled to support it was generated using the very measures that I have shown to have measurement error in them. At this point, I could not discard this concern, nor make any claims about potential endogeneity in assessments about the state of the economy. And I do recognize that it is an important topic to address in future research specifically designed to look into this matter. Notwithstanding, concerns on endogeneity have been empirically addressed - specifically when correcting the first source of measurement error - in Michelitch et al. (2012).

In a way, these new questions give a fresh start to researchers concerned with economic voting. With them, we are in a position to look again at the old foundational debates and gain new perspectives as to whether voters are *prospective*, *retrospective*, *sociotropic*, or *pocketbook* centered, but this time with measurements that are consistent with the theory. Similarly, new contributions can be generated to the growing literature on endogeneity in

economic voting assessments, but using measurements that are neither endogenous by design, nor measured with error by design as well.

7 Conclusion

Over 30 years of research on economic voting based on survey data have used a single question that requires respondents to provide an assessment about the state of the economy, in broad terms. During this time, the literature has accumulated mixed results that are, for the most part, inconsistent over time and within elections. Researchers have taken many creative routes to try to reconcile this divergence and explain its origins, perhaps without much success.

I argued in this paper that researchers have been using questions that are ill-suited to test economic voting empirically. These questions pose two very distinct problems: they are not consistent with what the theory requires researchers to model in their empirical tests, and they are also confounding information that should otherwise be unambiguous.

In an ideal world, we would want to capture respondents' economic assessments with a battery of survey questions that condition on each candidate being in office, accompanied with a response scale that can disambiguate whether respondents have a good or bad assessment when they respond that the economy will remain the same as it is now.

In stark contrast, researchers have utilized single questions that inquire about the state of the economy, absent any conditioning on candidates being in office, and with a response scale that has a mid-category with an ambiguous meaning.

The use of these questions may have deep implications for economic voting research. It is hard to gauge the magnitude of these effects *a priori*, but it is likely that the divergence

in results is caused by the use of a measure that is plagued with measurement error built into the question itself. But what is certain, is that these questions are not appropriate for what the theory that shapes economic voting requires.

The survey experiments presented here show that it is possible to design different survey instruments that are both consistent with the theory and that can reduce ambiguity in the responses, which are better suited to investigate economic voting empirically. In other words, with these proposed instruments we can reduce measurement error starting at the survey design stage.

Notes

[1]A joint proposal with Josh Tucker and Kristin Michelitch was presented to the ANES to explicitly address this problem prior to the fielding of the 2008 survey. The proposal was accepted, and the experiment repeated in the 2012 survey.

[2]A note of caution is in place: the conditional questions in 2012 omitted the temporal priming phrase "What about the next 12 months?", which was included in the naïve version. *Stricto sensu*, this prevents a full comparison between both versions. Nevertheless, I decided to include the results for the 2012 experiment as well as means to illustrate the patterns produced by the conditional questions, which are consistent with the 2008 results.

[3]As crafted, this experiment may not be applicable to settings other than face-to-face interviews, as one would think that respondents would not react well to a telephone interviewer or a screen that repeats the exact same question after it has been answered. Whether research designs should be "portable" across mediums is still not a settled matter in the survey research methods literature. Good arguments and evidence are marshaled in support of *unimode* or *multimode* methods (Dillman, Smyth & Christian 2008).

[4]Another interesting feature is that these last two questions seem to produce quite similar distributions, and incidentally belong to the two frontrunners, but there is not enough evidence to reach any conclusions from it. Previous experiments with theses questions conducted over 2008 - and not shown here - render clearly differentiated response distributions for the sitting president, and the other two competing candidates in the 2006 election. So, it is hard to reach any conclusions on this matter just yet.

[5]It is important to remember that both sets of responses were provided by exactly the same respondents. That is, the top panel shows responses to the mid-category in the initial 5-category scale, while the bottom one shows the answers to the follow-up 2-category question.

[6]Yet, in every case, the second largest concentration of votes is for EPN with the disambiguated category, which is not true with the ambiguous category where EPN would be the third largest concentration for those who thought the economy would be the same under EPN.

[7]Evidently, for those cases where no sitting President is running for reelection a slight modification is required in the conditioning of the questions: conditioning on all candidates would be made using the "challenger" phrasing of the question.

References

Alvarez, R. Michael & Jonathan Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* 39(3):714–744.

Alvarez, R. Michael & Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42(1):55–96.

Alvarez, R. Michael, Jonathan Nagler & Jennifer R. Willette. 2000. "Measuring the relative impact of issues and the economy in democratic elections." *Electoral Studies* 19(2-3):237253.

Alvarez, R. Michael, Jonathan Nagler & Shaun Bowler. 2000. "Issues, Economics, and the Dynamics of Multiparty Elections: The British 1987 General Election." *American Political Science Review* 94(1):131–149.

Anderson, Christopher J. 2007. "The End of Economic Voting? Contingency Dilemmas and the Limits of Democratic Accountability." *Annual Review of Political Science* 10:271–296.

Anderson, Christopher J., Silvia M. Mendes & Yuliya V. Tverdova. 2004. "Endogenous Economic Voting: Evidence from the 1997 British Election." *Electoral Studies* 23(4):687–708.

Bautista, René. 2012. An Overlooked Approach in Survey Research: Total Survey Error. In *Handbook of Survey Methodology for the Social Sciences*, ed. Lior Gideon. New York, NY: Springer.

Campbell, Angus, Phillip E. Converse, Warren E. Miller & Donald E. Stokes. 1960. *The American Voter*. Unabridged ed. Chicago, IL: University of Chicago Press.

Chzhen, Kat, Geoffrey Evans & Mark Pickup. forthcoming. "When do Economic Perceptions Matter for Party Approval?" *Political Behavior* .

Coombs, Clyde H. & Lolagene C. Coombs. 1976. "'Don't Know': Item Ambiguity or Respondent Uncertainty?" *Public Opinion Quarterly* 40(4):497–514.

Dillman, Don A., Jolene D. Smyth & Leah Melani Christian. 2008. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Third ed. ed. Hoboken, NJ: Wiley & Sons.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York, NY: Harper-Collins.

Erikson, Robert S. 1989. "Economic Conditions and The Presidential Vote." *American Political Science Review* 83(2):567–573.

Erikson, Robert S., Michael B. MacKuen, & James A. Stimson. 2000. "Bankers or Peasants Revisited: Economic Expectations and Presidential Approval." *Electoral Studies* 19(2-3):295–312.

Evans, Geoffrey & Mark Pickup. 2010. "Reversing the Causal Arrow: The Political Conditioning of Economic Perceptions in the 2000-2004 U.S. Presidential Election Cycle." *Journal of Politics* 72(4):1236–1251.

Evans, Geoffrey & Robert Andersen. 2006. "The Political Conditioning of Economic Perceptions." *Journal of Politics* 68(1):194–207.

Fair, Ray C. 1982. "The effects of economic events on votes for president." *Review of Economics and Statistics* 60(2):159–173.

Fiorina, Morris P. 1981. *Retrospective voting in American national elections*. New Haven, CT: Yale University Press.

Fricker, Scott, Mirta Galesic, Roger Tourangeau & Ting Yan. 2005. "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69(3):370–392.

Gelman, Andrew, Gary King & Chuanhai Liu. 1998. "Not Asked and Not Answered: Multiple Imputation for Multiple Surveys." *Journal of the American Statistical Association* 93(443):846–857.

Gerber, Alan S. & Gregory A. Huber. 2010. "Partisanship, Political Control, and Economic Assessments." *American Journal of Political Science* 54(1):153–173.

Glasgow, Garrett & Roberto A. Weber. 2005. "Is there a relationship between election outcomes and perceptions of personal economic well-being? A test using post-election economic expectations." *Electoral Studies* 24(4):581–601.

Glucksberg, Sam & Michael McCloskey. 1981. "Decisions about Ignorance: Knowing that you don't know." *Journal of Experimental Psychology: Human Learning and Memory* 7(5):311–325.

Goodman, Saul & Gerald H. Kramer. 1975. "Comment on Arcelus and Meltzer, the effect of aggregate economic conditions on congressional elections." *American Political Science Review* 69(4):1255–1265.

Hausman, Jerry A. & David A. Wise. 1978. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences." *Econometrica* 46(2):403–427.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8(1):83–98.

Hibbs, Douglas A. 1987. *The American political economy: Macroeconomics and electoral politics in the United States*. Cambridge, MA: Harvard University Press.

Hopkins, Daniel J. & Gary King. 2010. "Improving Anchoring Vignettes: DesDesign Surveys to Correct Interpersonal Incomparability." *Public Opinion Quarterly* 74(2):201–222.

Huber, Gregory A. & Celia Paris. 2013. "Assessing The Programmatic Equivalence Assumption in Question Wording Experiments: Understanding Why AmAmerica Like Assistance to the Poor More than Welfare." *Public Opinion Quarterly* 77(1):385–397.

Kiewiet, D. Roderik. 1983. *Macroeconomics and Micropolitics*. Chicago, IL: University of Chicago Press.

Kiewiet, D. Roderik & Douglas Rivers. 1985. A retrospective on retrospective voting. In *Economic Conditions and Electoral Outcomes: The United States and Western Europe*, ed. Eulau Heinz & Michael S. Lewis-Beck. New York, NY: Algora Publishing.

Kinder, Donald R. & D. Roderick Kiewiet. 1981. "Sociotropic Politics: The American Case." *British Journal of Political Science* 11(2):129–161.

King, Gary, James Honaker, Ann Joseph & Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An alternative algorithm for multiple imputation." *American Political Science Review* 95(1):49–69.

Kramer, Gerald H. 1971. "Short-Term Fluctuations in U.S. Voting Behavior: 1896-1964." *American Political Science Review* 65(1):131–143.

Krause, George A. 1997. "Voters, Information Heterogeneity, and the Dynamics of Aggregate Economic Expectations." *American Journal of Political Science* 41(4):1170–1200.

Kroh, Martin. 2007. "Measuring Left-Right Political Orientation: The Choice of Response Format." *Public Opinion Quarterly* 71(2):204–220.

Krosnick, Jon A. 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys." *Applied Cognitive Psychology* 5(3):213–236.

Krosnick, Jon A. & Stanley Presser. 2010. Question and Questionnaire Design. In *Handbook of Survey Research*, ed. Peter V. Marsden & James D. Wright. New York, NY: Emerald.

Ladner, Matthew & Christopher Wlezien. 2007. "Partisan Preferences, Electoral Prospects, and Economic Expectations." *Comparative Political Studies* 40(5):571–596.

Lanoue, David J. 1994. "Retrospective and Prospective Voting in Presidential-Year Elections." *Political Research Quarterly* 47(1):193–205.

Lewis-Beck, Michael S. 1988. "Economics and the American Voter: Past, Present, Future." *Political Behavior* 10(1):5–21.

Lewis-Beck, Michael S. & Mary Stegmaier. 2000. "Economic Determinants of Electoral Outcomes." *Annual Review of Political Science* 3:183–219.

Lewis-Beck, Michael S. & Tom Rice. 1992. *Forecasting Election*. Washington, DC: Congressional Quarterly Press.

Linn, Suzanna, Jonathan Nagler & Marco A. Morales. 2010. Economics, Elections, and Voting Behavior. In *The Oxford Handbook of American Elections and Political Behavior*, ed. Jan E. Leighley. New York, NY: Oxford University Press.

Lockerbie, Brad. 1991. "Prospective Economic Voting in U. S. House Elections, 1956-88." *Legislative Studies Quarterly* 16(2):239–261.

MacKuen, Michael B., Robert S. Erikson & James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *American Political Science Review* 86(3):597–611.

Markus, Gregory. 1988. "The Impact of Personal and National Economic Conditions on the Presidential Vote: A Pooled Cross-Sectional Analysis." *American Journal of Political Science* 32(1):137–154.

Michelitch, Kristin, Marco A. Morales, Andrew Owen & Joshua Tucker. 2012. "Looking to the future: Prospective Economic Voting in 2008 Presidential Elections." *Electoral Studies* 31(4):838–851.

Nickelsburg, Michael & Helmut Norpoth. 2000. "Commander-in-chief or chief economist? The president in the eye of the public." *Electoral Studies* 19(2-3):313–323.

Norman, Donald A. 1973. Memory, Knowledge, and the Answering of Questions. In *Contemporary Issues in Cognitive Psychology, The Loyola Simposium*, ed. Robert L. Solso. Washington, DC: Winston.

Palmer, Harvey D. & Raymond M. Duch. 2001. "Do Surveys Provide Representative or Whimsical Assessments of the Economy?" *Political Analysis* 9(1):58–77.

Pickup, Mark & Geoffrey Evans. 2013. "Addressing the Endogeneity of Economic Evaluations in Models of Political Choice." *Public Opinion Quarterly* 77(3):735–754.

Reiter, Jerome P., Trivellore E. Raghunathan & Satkartar K. Kinney. 2006. "The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data." *Survey Methodology* 32(2):143–149.

Rubin, Donald B. 1976. "Inference and Missing data." *Biometrika* 63(3):581–592.

Rubin, Donald B. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91(434):473–489.

Schwarz, Norbert. 1990. "What Respondents Learn from Scales: The Informative Functions of Response Aletrantives." *International Journal of Public Opinion Research* 2(3):274–285.

Schwarz, Norbert. 1995. "The Numeric Values of Rating Scales: A Comparison of their Impact in Mail Surveys and Telephone Interviews." *International Journal of Public Opinion Research* 7(1):72–74.

Schwarz, Norbert, Bärbel Knäuper, Hans J. Hippler, Elisabeth Noelle-Neuman & Leslie Clark. 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55(4):570–582.

Schwarz, Norbert, Carla E. Grayson & Bärbel Knäuper. 1998. "Formal Features of Rating Scales and the Interpretation of Question Meaning." *International Journal of Public Opinion Research* 10(2):177–183.

Tilley, James & Sara Hobolt. 2011. "Is the Government to Blame? An Experimental Test of How Partisanship Shapes Perceptions of Performance and Responsibility." *Journal of Politics* 73(2):316–330.

Tourangeau, Roger, Lance J. Rips & Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press.

Tufte, Edward R. 1975. "Determinants of the Outcomes of Midterm Congressional Elections." *American Political Science Review* 69(3):812–826.

van der Eijk, Cees, Mark Franklin, Froukje Demantc & Wouter van der Brugd. 2007. "The Endogenous Economy: 'Real' Economic Conditions, Subjective Economic Evaluations and Government Support." *Acta Politica* 42(1):1–22.

Visser, Penny S., Jon A. Krosnick, Jesse F. Marquette & Michael F. Curtin. 2000. Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters and Response Order Effects. In *Election Polls, the News Media, and Democracy*, ed. Paul L. Lavrakas & Michael Traugott. New York, NY: Chatam House.

Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A new Guide to the New Science of Survey Research*. Chicago, IL: University of Chicago Press.

Whitten, Guy D. & Harvey D. Palmer. 1999. "The Electoral Impact of Unexpected Inflation and Economic Growth." *British Journal of Political Science* 29(4):623–639.

Wilcox, Nathaniel & Christopher Wlezien. 1993. "The Contamination of Responses to Survey Items: Economic Perceptions and Political Judgments." *Political Analysis* 5(1):181–213.

Wlezien, Christopher, Mark Franklin & Daniel Twiggs. 1997. "Economic Perceptions and Vote Choice: Disentangling the Endogeneity." *Political Behavior* 19(1):7–17.

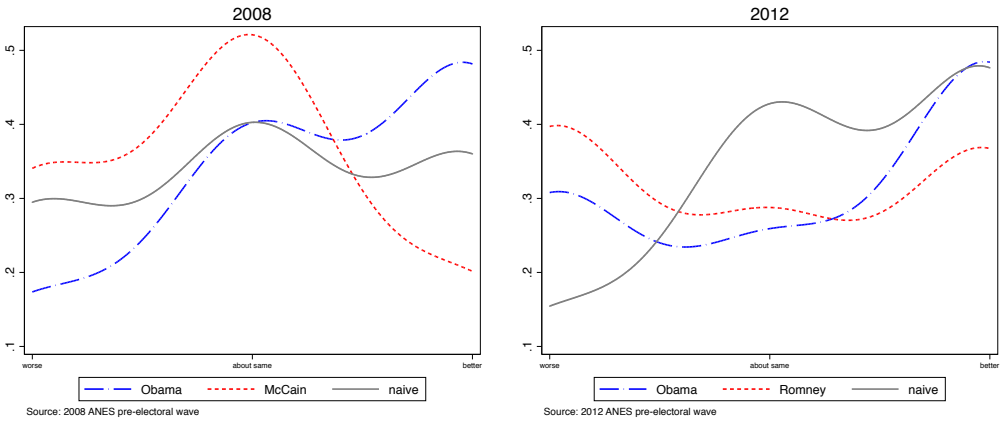

Figure 1: Response distributions for prospective economic evaluation questions fielded on the 2008 and 2012 ANES pre-electoral waves. Gray line shows kernel density for response distributions on the standard (naïve) question used in election studies. Blue and red dotted lines show kernel density response distribution to prospective economic evaluation questions that condition on the Democratic (Barack Obama) or Republican (John McCain/Mitt Romney) candidates being in office.

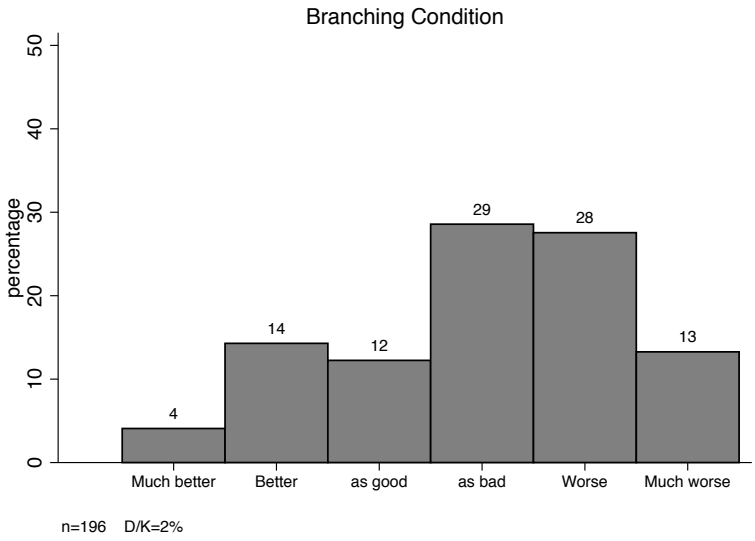

Figure 2: Response distribution for questions fielded under the " branching condition" on the September 2008 DEFOE survey. When a "same" response is volunteered, respondents are asked to clarify whether that means "just as good" or "just as bad" as it is today.

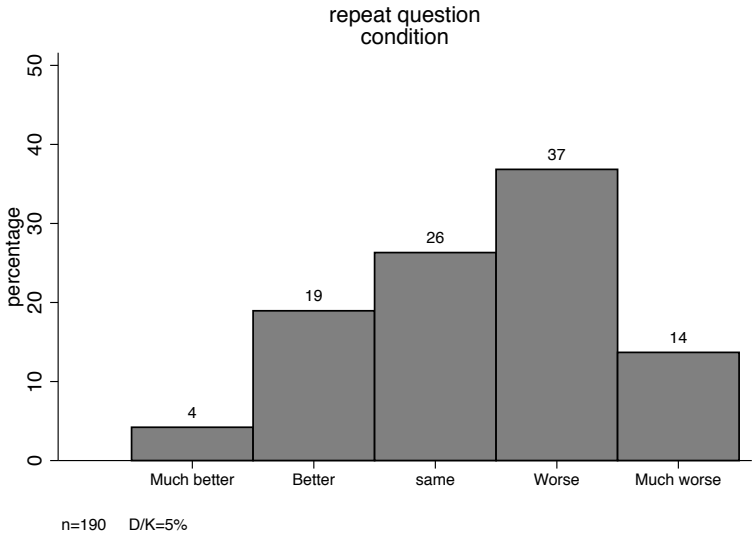

Figure 3: Response distribution for questions fielded under the "repeat question condition" on the September 2008 DEFOE survey. When a "same" response is volunteered, it will only be coded as such if respondents volunteer this answer on three consecutive times.

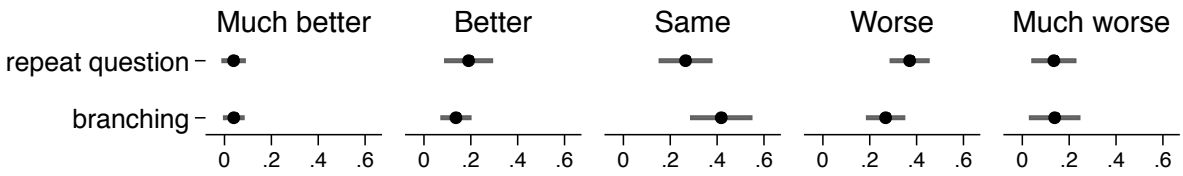

Figure 4: Comparison of response distributions for both experimental conditions, under a 5-category distribution, where the mid-categories in the "branching condition" are collapsed into a single category. Dots and whiskers are point estimates and their associated 95% confidence intervals. Estimates performed on the September 2008 DEFOE survey.

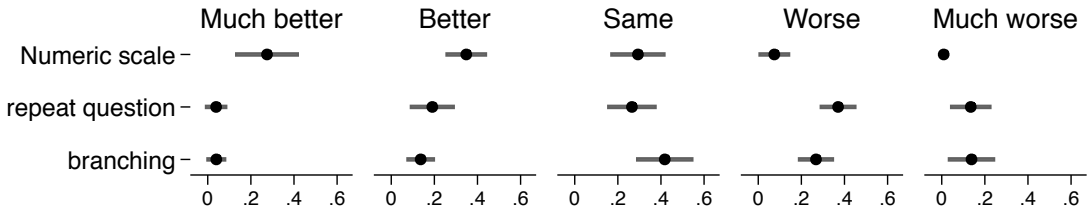

Figure 5: Comparison of response distributions for both experimental conditions, under a 5-category distribution, where the mid-categories in the "branching condition" are collapsed into a single category. Distribution for "numeric condition" results from a numeric response scale with 1 labeled "much worse" and 5 labeled "much better". Dots and whiskers are point estimates and their associated 95% confidence intervals. Estimates performed on the September 2008 DEFOE survey.

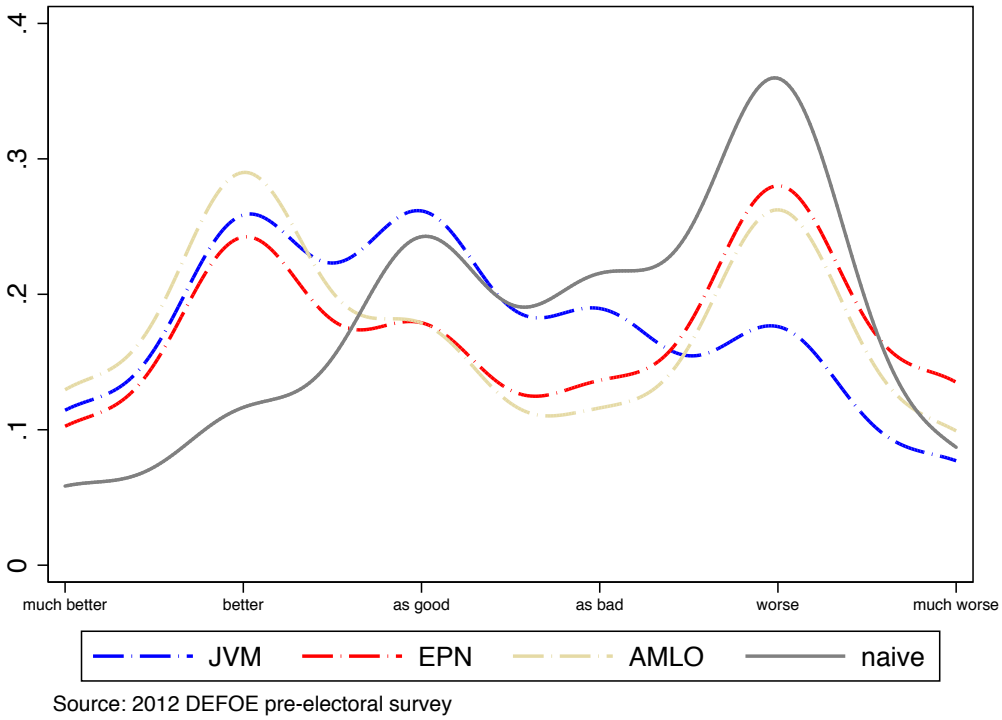

Figure 6: Response distributions for split-sample experiments on prospective economic evaluation questions fielded on the 2012 DEFOE pre-electoral survey. Gray line shows kernel density for response distributions on the standard (naïve) question used in election studies. Blue, red, and yellow dotted lines show kernel density response distribution to prospective economic evaluation questions that condition on Josefina Vázquez Mota (JVM), Enrique Peña Nieto (EPN) or Andrés Manuel López Obrador (AMLO) being in office.

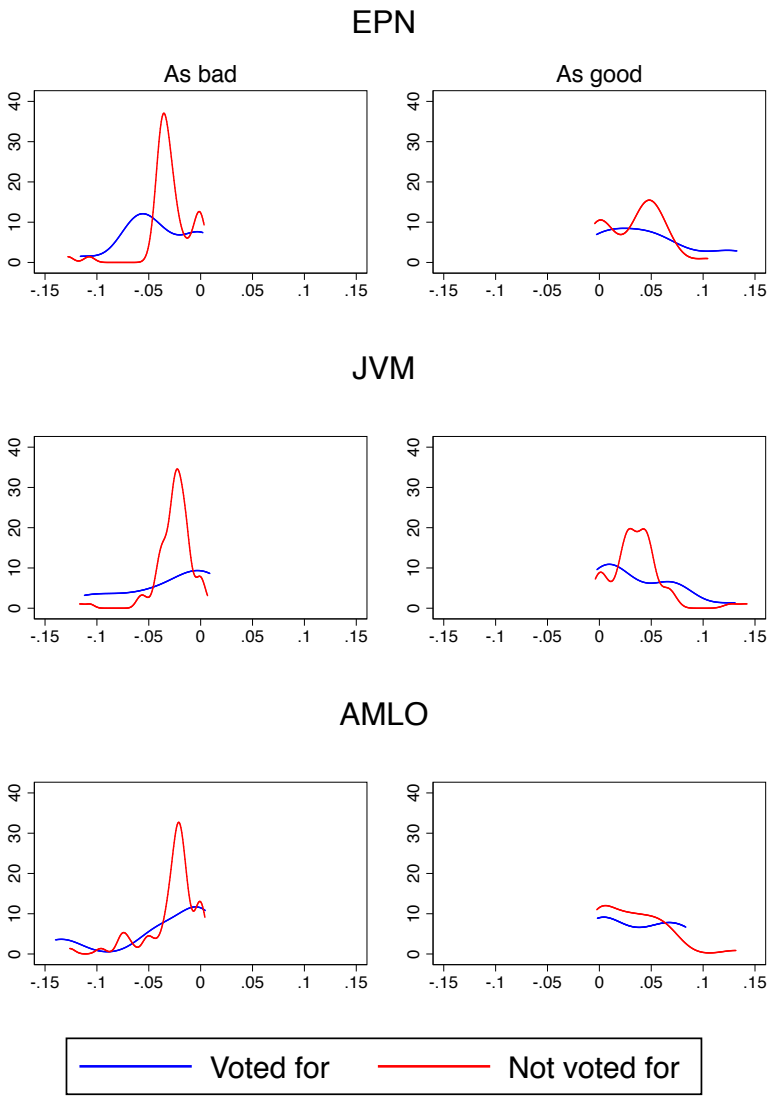

Figure 7: Distribution of changes in the probability of selecting the indicated (row) candidates given that a respondent provided the indicated (column) assessment between a model that estimates vote choice using the disambiguated question and another using the ambiguous question. Blue lines indicates distribution of probability changes for voters who selected the indicated (row) candidate. Red line indicates distribution of probability changes for voters who did not selecte the indicated (row) candidate.

Table 1: Response distributions for split-sample experiments on prospective economic evaluation questions fielded on the 2008 ANES pre-electoral wave. Numbers in cells are percentages (of valid responses).

		Conditional	
Assessment	**Naïve**	**Obama**	**McCain**
better	34.48	46.75	18.13
same	37.58	37.40	49.91
worse	27.94	15.86	31.96
n (valid)	1,163	1,091	1,092
n (total)	1,193	1,129	1,129

Table 2: Response distributions for prospective economic evaluation questions fielded on the 2012 ANES pre-electoral wave. Numbers in cells are percentages (of valid responses).

		Conditional	
Assessment	**Naïve**	**Obama**	**Romney**
better	45.99	47.54	38.70
same	40.30	22.58	25.58
worse	13.71	29.88	35.72
n (valid)	5,777	5,867	5,793
n (total)	5,916	5,916	5,916

Table 3: Response distributions for split-sample experiments on prospective economic evaluation questions fielded on the 2012 DEFOE pre-electoral survey. Numbers in cells are percentages (of valid responses).

		Conditional		
Assessment	**Naïve**	**JVM**	**EPN**	**AMLO**
much better	7.19	7.00	12.36	8.82
better	34.89	16.60	27.03	25.49
as good	19.06	17.20	11.78	9.80
as bad	23.02	24.40	16.41	16.27
worse	10.43	24.40	23.17	27.84
much worse	5.40	10.40	9.27	11.76
n (valid)	556	500	518	510
n (total)	600	600	600	600

Table 4: Response distributions for prospective economic assessment questions conditional on each of the three candidates being in office, classified by vote intention as per 2012 DEFOE pre-electoral survey. Numbers in cells are (column) percentages.

	Vote Intent EPN			**Vote Intent JVM**			**Vote Intent AMLO**		
	Conditional			**Conditional**			**Conditional**		
Assessment	**EPN**	**JVM**	**AMLO**	**EPN**	**JVM**	**AMLO**	**EPN**	**JVM**	**AMLO**
much better	**27.86**	2.66	4.35	1.90	**20.18**	1.96	3.45	3.52	**18.99**
better	**48.26**	8.51	8.15	14.29	**48.62**	11.76	11.72	5.63	**60.13**
as good	8.46	18.62	10.33	20.00	14.68	13.73	9.66	16.20	5.70
as bad	7.96	27.66	20.11	13.33	9.17	21.57	28.28	30.28	6.96
worse	4.98	**30.32**	**36.96**	**40.95**	4.59	**37.25**	**31.03**	**30.28**	5.06
much worse	2.49	**12.23**	**20.11**	**9.52**	2.75	**13.73**	**15.86**	**14.08**	3.16
total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 5: Response distributions for mid-categories of prospective economic assessments conditional on each one of the candidates being in office, as segmented by vote intention. Each row corresponds to the percentage of respondents that manifest an intent to vote for the specified candidates. Top part of the table corresponds to responses to disambiguated scale. Bottom part to responses to ambiguous scale. Data corresponds to 2012 DEFOE pre-electoral survey.

Conditional Assessment		**Vote Intent**			
		EPN	**JVM**	**AMLO**	Total
EPN	as good	32.7	**40.4**	26.9	100
	as bad	22.5	19.7	**57.8**	100
JVM	as good	**47.3**	21.6	31.1	100
	as bad	**49.5**	9.5	**41.0**	100
AMLO	as good	**45.2**	33.3	21.5	100
	as bad	**52.9**	31.4	15.7	100
EPN	same	26.8	28.4	**44.8**	100
JVM	same	**48.6**	14.5	36.9	100
AMLO	same	**50.0**	32.2	17.8	100

Table 6: Classification statistics produced by vote-choice model that utilize disambiguated and ambiguous conditional prospective economic evaluations. Estimates performed on 2012 DEFOE pre-electoral survey. PCP is "Percentage Correctly Predicted" and ePCP is Herron's (1999) "expected Percentage Correctly Predicted".

	Predicted (disambiguated)			**Predicted (ambiguous)**			
Observed	**EPN**	**JVM**	**AMLO**	**EPN**	**JVM**	**AMLO**	**n (total)**
EPN	384	19	32	386	18	31	435
JVM	77	175	13	75	172	18	265
AMLO	52	12	238	63	11	235	309
n (total)	520	206	283	524	201	284	1,009
		PCP	78.98%		**PCP**	78.59%	
		ePCP	62.65%		**ePCP**	62.66%	

Appendix

A Experiment 1: Correcting question phrasing

The questions were fielded using a split-sample design in the 2008 pre-electoral survey, whereby the full ANES sample (n=2,322) was randomly divided in two subsamples. One of them received the naïve question, and the other one a battery of two questions corresponding to each one of the two main candidates competing in the presidential election. In 2012 pre-electoral survey, the naïve and the conditional questions were administered to the full sample of respondents (n=5,916), although in different parts of the questionnaire.

A.1 2008 question phrasing

What about the next 12 months? If [Barack Obama/John McCain] wins the election, do you expect the economy, in the country as a whole, to get BETTER, stay ABOUT THE SAME, or get WORSE?

A.2 2012 question phrasing

If [Barack Obama/Mitt Romney] wins the election, do you expect the economy, in the country as a whole, to get BETTER, stay ABOUT THE SAME, or get WORSE?

B Experiment 2: Correcting response scale

Mexican polling firm DEFOE fielded a national probability sample survey on September 2008 (n=800), on the general population over 18 years of age. The survey included a split-sample design, with four experimental conditions. Each condition was randomly assigned to 200 respondents. (The fourth condition is not reported here). The exact phrasing of each one of the relevant experimental conditions is as follows:

B.1 Branching Condition

Thinking about the economy in the country for the coming year, would you say that the economy will be much better, better, about the same, worse, or much worse than it currently is? (INTERVIEWER: if respondent answers same, insisit just as good or just as bad)

B.2 Repeat Question Condition

Thinking about the economy in the country for the coming year, would you say that the economy will be much better, better, about the same, worse, or much worse than it currently is? (INTERVIEWER: if respondent answers same, repeat the question. If after three repetitions still responds same, code as "Same")

B.3 Numeric Scale Condition

Thinking in a scale from 1 to 5 where 1 would be "much worse than it is now" and 5 would be "much better than it is now", how would you rate the economic situation in the country a year from today? (RESPONDENTS RECEIVE A CARD WITH A SCALE REPRESENTING THIS INFORMATION)

C Experiment 3: Conditional Questions with Unambiguous Response Scales

As part of its ongoing 2012 Election Panel survey, the Mexican polling firm DEFOE fielded a national probability sample survey on June 2012 (n=1,200). The survey included a split-sample design, with two experimental conditions. Each condition was randomly assigned to 600 respondents. The exact phrasing of each one of the experimental conditions is as follows:

C.1 Conditional Questions

Thinking about the economy in the country for the coming year, would you say that if [PAN's Josefina Vázquez Mota/PRI's Enrique Peña Nieto/PRD's Andrés Manuel López Obrador] were elected President, the economy will be much better, better, about the same, worse, or much worse than it currently is? (INTERVIEWER: if respondent answers same, insist just as good or just as bad)

C.2 Naïve Question

Thinking about the economy in the country for the coming year, would you say that the economy will be much better, better, about the same, worse, or much worse than it currently is? (INTERVIEWER: if respondent answers same, insist just as good or just as bad)

D 2012 Mexican presidential election vote choice model estimates

Two vote choice models were estimated on the 2012 DEFOE pre-electoral survey data to test whether different estimates are produced when using the ambiguous or the disambiguated responses to the conditional prospective economic assessment questions proposed in this paper.

To make a better use of the available data, and given that the conditional questions were asked only on half of the sample, the dataset was multiply imputed. The survey experiment design makes it perfectly suitable to perform Multilple Imputation on the dataset, since *ignorability* in the missing data mechanism holds as a result of the random assignment of the experimental questions (Rubin 1976, Gelman, King & Liu 1998). Hence, the model can be estimated using the full sample, as opposed to just half of it.

In addition of using all variables that are included in the model, as well as those that can assist to better estimate the covariance matrix common to all respondents, the imputation model accounts for the sample design to avoid inconsistent estimates (Rubin 1996, King et al. 2001, Reiter, Raghunathan & Kinney 2006) by using "cluster effects" for each surveyed precinct.

The vote choice model was estimated using a multinomial probit model (Hausman & Wise 1978) which allows researchers to overcome the IIA assumption that must be made when estimating other types of models, and which is most likely inadequate in mutiple candidate election settings (Alvarez & Nagler 1995, 1998). In particular, we allow the covariance matrix in the model's random component (Σ) to have non-zero correlations to be estimated. To identify the estimation, disturbances are assumed to be homoskedastic ($\sigma^2_{PRI} = \sigma^2_{PRD} = \sigma^2_{PAN} = 1$) and the correlation between the PRI and PRD disturbances is assumed to be zero ($\sigma_{PRI,PRD} = 0$). Hence, the covariance matrix to be estimated is:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & & \\ \sigma_{PRI,PAN} & 1 & \\ 0 & \sigma_{PAN,PRD} & 1 \end{bmatrix} \tag{1}$$

The analysis is performed with $n = 1{,}099$ cases, since we discard cases that would vote for small-party or non-registered candidates. Point estimates were computed as per the "Rubin rules" (Rubin 1976, 1996).

Table 7: Multinomial Probit estimates for Presidential election, using Conditional Prospective Economic Assessments (disambiguated)

	PRI/PAN	**PRD/PAN**
Cond Econ Assessment (much better)	1.976*** (0.247)	
Cond Econ Assessment (better)	1.421*** (0.141)	
Cond Econ Assessment (as good)	0.474*** (0.131)	
Cond Econ Assessment (worse)	-0.576*** (0.151)	
Cond Econ Assessment (much worse)	-0.874*** (0.254)	
Male	0.265 (0.162)	0.490*** (0.135)
Age	-0.006 (0.006)	0.001 (0.004)
Income	-0.032 (0.099)	0.087 (0.067)
Education	-0.199** (0.085)	-0.136** (0.066)
Intercept	1.078*** (0.102)	-0.033 (0.083)
$\sigma_{PRI,PAN}$	-0.263 (0.230)	
$\sigma_{PRD,PAN}$	0.241* (0.131)	
Log-Likelihood	-730.322	
LR-test	$\chi^2_{[15]}$=244.279***	
n	1,009	
MI sets	10	

Significance: 1% *** / 5% ** / 10%* two-tailed.

Table 8: Multinomial Probit estimates for Presidential election, using Conditional Prospective Economic Assessments (ambiguous)

	PRI/PAN	**PRD/PAN**
Cond Econ Assessment	1.753***	
(much better)	(0.224)	
Cond Econ Assessment	1.202***	
(better)	(0.145)	
Cond Econ Assessment	-0.806***	
(worse)	(0.150)	
Cond Econ Assessment	-1.106***	
(much worse)	(0.259)	
Male	0.270*	0.488***
	(0.162)	(0.134)
Age	-0.007	0.001
	(0.006)	(0.004)
Income	-0.026	0.094
	(0.098)	(0.066)
Education	-0.216**	-0.150**
	(0.086)	(0.065)
Intercept	1.155***	0.030
	(0.103)	(0.084)
$\sigma_{PRI,PAN}$	-0.278	
	(0.228)	
$\sigma_{PRD,PAN}$	0.230*	
	(0.132)	
Log-Likelihood	-738.940	
LR-test	$\chi^2_{[14]}$=237.791***	
n	1,009	
MI sets	10	

Significance: 1% *** / 5% ** / 10%* two-tailed.