Designing and Fielding Surveys in Labor Economics*

Sydnee Caldwell

First Version: Fall 2020, last updated Spring 2025

These lecture notes were developed for my PhD class at UC Berkeley (Econ 250B). They are intended as a practical guide for Berkeley PhD students interested in designing and fielding original survey instruments. Please contact me if you have any comments (or if you find them useful!)

1 Surveys: Introduction

Economists have long used surveys to inform our understanding of the labor market. Famous examples (which I highly recommend) include Truman Bewley's "Why Wages Don't Fall" and Lloyd Reynolds' "The Structure of the Labor Market".

There are several econometric issues to consider when using *existing* survey data—whether to use sampling weights, etc. But economists today are increasingly *running* their own surveys (or survey modules) to collect data. This lecture covers common issues that arise. Some of the issues parallel those in the experimental literature. I highly recommend Stefanie Stantcheva's overview of surveys, which goes over many details beyond the scope of this lecture. If you are serious about running surveys, you should go beyond the economics literature, however. I am happy to recommend specific textbooks and other resources.

Surveys have recently become much more common in "top 5" economics journals, as shown by Dutz et al. (2021). So even if you are not planning on running a survey yourself, you may need to learn how to evaluate such research.

1.1 Why Run a Survey?

There are several reasons to run your own survey: collecting baseline data for experiments, measuring outcomes missing from administrative data (e.g., rejected job offers, worker beliefs), or embedding experiments directly. However, collecting original data usually requires

*Thanks to Kyle Neary for outstanding help revising these notes.

Figure 1: Use of survey data in top-five publications

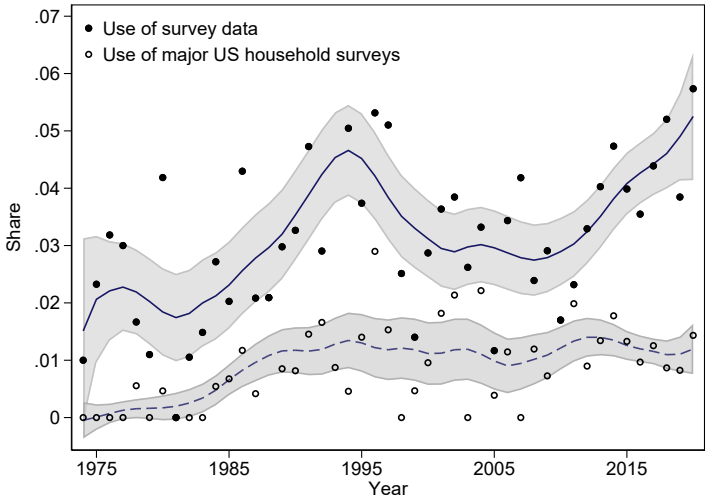


Figure 1: Dutz et al. (2021)

significant up-front investment and careful planning—you often can't revise your data collection afterward. Before going to the field, clearly outline the paper, write a draft questionnaire or experimental design, develop a simple theoretical model (even if you don't plan to include one in the paper), and prepare blank tables for anticipated results.

Think through possible outcomes: are all patterns of results interpretable and publishable? Can your design distinguish between different theoretical mechanisms? If not, adjust your data collection strategy accordingly. I usually design projects such that all plausible outcomes are informative and interesting; this reduces some the risk, but tastes vary.

You should take all of these ingredients—the model and blank tables—and put them together in a slide deck. Add additional slides for the introduction/motivation and for explaining how you would interpret different combinations of results. Use this deck to pitch the project to other students (and to me!). Once you run the pilot, you can fill in this deck with preliminary results.

1.2 Examples for Lecture

Throughout the lecture, I will refer back to the Bloom and Van Reenen (2007) paper discussed in a previous lecture. Since the original paper, the authors have expanded the scope of the survey considerably, adding additional industries and countries.

I discuss this survey because the authors designed pretty much all aspects of it themselves. They identified a sampling frame based on public use data, trained interviewers, fielded the

survey, and validated the responses. They then used the survey in a number of papers, some of which involved linking the responses to administrative data sources. The WMS team also provide a wealth of information on what they did on their website.

I will also refer quite frequently to Stefano DellaVigna, Jörg Heining, Johannes F Schmieder and Simon Trenkle (2022) and to the firm- and worker-level surveys I developed in Germany. I have not gone over DHST in class, but the survey was very carefully designed. This has the advantage of being a survey of individuals, not firms (WMS is a firm survey). I'll talk a bit about designing surveys of both sides of the market; this is what I did in Germany.

2 Surveys: Field Logistics

2.1 Identifying Potential Respondents

The first challenge in running a survey (unless coincident with an experiment) is selecting the sampling frame. If you are working with a government agency or firm, it may be easy to construct a sampling frame that is representative of the population of interest. For instance, you may select individuals for inclusion in your survey from an administrative database or from a roster of employees (potentially subject to some inclusion criteria).

Sometimes, however, people want to run a survey and do not know how to contact the population of interest. For instance, you may want to conduct a survey of full-time employed individuals in the United States. If this is the case, there are a few commonly used options (this is non-exhaustive):

Online Platforms First, you can use an online platform such as mTurk or Prolific. The advantage of these platforms is that they are easy to implement and allow for fairly sophisticated filtering. You can put a survey online and get responses in $<$1 day. The platforms are also generally low cost—you pay a fee to the platform and pay participants based on the expected completion time. These platforms typically recommend payment slightly above minimum wage, e.g., ~$3 per 10-minute survey. Of course, one thing you should ask yourself is how representative individuals who work on mTurk or Prolific—who spend their time apart from leisure or their regular job completing surveys—are likely to be.

There is an ongoing discussion of external validity in surveys and whether these platforms generate representative samples (see, e.g., Dutz et al., 2021). Due to concerns regarding external validity, these platforms may be most useful as sampling frames for projects where there is an experimental component (so you are assured of internal validity), where the primary contribution is econometric, or where you are trying to reach individuals for whom the

modest financial incentives on the platform are likely to be meaningful. I would be skeptical of a survey of CEOs or human resource professionals fielded on such a platform—why would someone with one of these jobs enroll in a survey, especially for such low compensation?

Even if you don't use one of these platforms for your survey, I encourage you to look at them as a way to pre-test or pilot your questions. While you should always get feedback on your questions from your advisors and classmates, you should also realize that these populations may not be very similar to your potential respondent pool! You may get better insights into question design if you contact individuals through these online platforms who look and think in ways that are similar to your potential pool. Because the respondents on this platform do many surveys, they can often provide valuable feedback on question wording based on their recent experience with other surveys. If you are trying to target a specific population, do in-depth interviews with members of that population when developing your questions.

Commercially Recruited Panels Sometimes researchers (usually groups of more senior researchers) run their own surveys using survey research firms to recruit participants. The advantage of doing this is that survey research firms offer skilled interviewers and multiple modes of contacting respondents. By working directly with the firm, it may be possible to produce representative samples and to implement strategies to deal with non-response. The disadvantage is that it is often very expensive. This makes sense: it is a lot of work to recruit individuals to a survey. Examples of such firms include Qualtrics, Bilendi, Prolific Academic, and Dynata. These firms vary in how much back-end support they vs. you provide. Generally the costs will be lower if you are doing more of the work.

One thing to note is the distinction between probability and non-probability samples. A probability sample is what you think it is: a sample where you know the exact probability a given respondent was invited to participate. A non-probability sample is one where you do not know the exact probability an individual would be sampled. Many commercial survey platforms use non-probability sampling. A common technique is quota sampling. In this type of sample, the company aims to match certain characteristics of the population. Individuals are then sampled until quotas are met.

Adding Questions to Existing Surveys Embedding your survey questions in an existing survey reduces the amount of work *you* have to do because you aren't actually fielding the survey yourself. This generally means you are not involved in recruiting the participants, designing strategies to deal with non-response, developing the recruitment language, programming the questionnaire, etc. These tasks are not easy and take a large amount of

time, especially if you want to do them properly.

Embedding your questions in an existing survey dramatically reduces costs. In part this is because the hard work of recruiting the panel has already been done—and the costs are shared among all individuals including questions on the survey. There are a few panels where you can essentially pay or apply to have your questions added:

1. German Socio-Economic Panel (GSOEP) (see: https://www.diw.de/en/diw_01.c.601784.en/soep-is_innovative_modules.html for more information).

2. NORC Amerispeak panel (see: https://amerispeak.norc.org/us/en/amerispeak/about-amerispeak/overview.html for more information).

3. World Management Survey (WMS) (see: https://worldmanagementsurvey.org/data/collect-new-data/run-your-survey/ for more information).

This is a non-exhaustive list (e-mail me if you think something should be added). A recent example of a paper that used this approach is Simon Jäger, Christopher Roth, Nina Roussille and Benjamin Schoefer (2024). The GSOEP regularly has open calls to add questions to the survey. Anyone can submit questions! A subset of the responses (for current waves it seems ~550 individuals) can be linked to administrative social security data.

Organization-Provided Sampling Frames Firms or organizations may also run surveys for you. For instance, a firm may mail a questionnaire to employees on your behalf. Or you may survey individuals who show up or receive a particular service (e.g., individuals who show up at a tax prep site or who use a particular job search program). Examples of the former approach include Handel and Kolstad (2015)'s investigation of frictions in insurance plan choice (which you will learn about if you take IO).

Another common approach these days is to work with professional organizations, such as groups of human resource professionals, or the like. A recent example is the survey in the paper by Zoe B Cullen, Shengwu Li and Ricardo Perez-Truglia (2022) on pay benchmarking. This can be a good approach if you want to identify individuals within a certain profession and if you are able to get a reputable organization to field your survey.

One issue is that response rates tend to be lower at organizations that send lots of surveys, the organizations most likely to say "yes" to fielding your survey. Organizations typically won't provide you with the information necessary to gauge selection into response. These surveys are often most useful to provide corroborating evidence for your main analysis (as is the case in both papers mentioned here).

Public Use or Administrative Datasets By far the hardest approach is to develop your own sampling frame using public use or administrative datasets. This is more difficult than other methods because you generally have to—beyond simply picking the sample—organize all details of the outreach and questionnaire yourself.

For instance, Bloom and Van Reenen (2007) used novel survey data on manufacturing firms in several countries. The sampling frame came from the public use Amadeus and Compustat datasets. Because of this, the authors were able to examine the characteristics of responding and non-responding firms. DHST used administrative data at the IAB—a group within the German Federal Employment Agency—to select individuals for inclusion in their survey of UI recipients. This is simpler than the Bloom and Van Reenen approach because the IAB already had contact information for each of the individuals in the database.

Going down this road can be costly (unless the cost is picked up by the statistical agency you are working with) and is definitely difficult. The best case scenario is that you field a survey with a statistical agency that has experience conducting these surveys and already has an established infrastructure. But there are often unanticipated issues that arise.

Other Methods People have gotten creative in recent years with other sampling frames. For instance, people have used Facebook ads to target individuals with specific political leanings or employment histories.

2.2 Evaluating your Sampling Frame

Sampling frame bias occurs when your sampling frame does not accurately represent the population of interest. A famous example of sampling bias comes from the 1948 presidential election. The Chicago Daily Tribute ran an election poll and went to press with the headline "DEWEY DEFEATS TRUMAN" (Truman won with a comfortable margin). The issue was that they conducted their survey by phone; wealthy people were more likely to have a phone and were also more likely to identify as Republican.

2.3 Fielding the Survey

Survey Delivery Methods and Recruitment

Once you have identified your respondents, you need to elicit your responses. Surveys can be delivered online, via phone (call/text), or in-person. Delivery method need not match contact method. For instance, you could invite participants via mail, but have them complete the survey online (using a QR code or URL). In some cases you may be able to use "mixed

methods" (i.e., combinations) to reach respondents. For example, you may offer individuals without text messaging the opportunity to respond in another way.

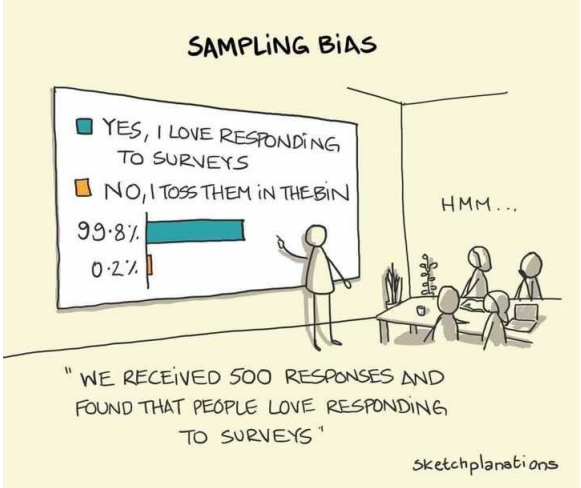
An importance difference is whether the interviewer interacts with the respondent. This can affect individuals' responses for a few reasons. First, individuals may be less willing to say things to an interviewer (vs. writing in an "anonymous" survey). Second, interviewers are able to clarify areas of potential confusion. And for open-ended questions, some interviewers may be more skilled at eliciting responses than others. If your paper heavily relies on open-ended questions and discussions between interviewer and interviewee, structure things so that you are able to take out interviewer fixed effects if necessary. Interviewers should also be trained to ensure consistency. The World Management Survey website has a great deal of information on what they did to ensure consistency (though they also often take out interviewer fixed effects).

Another distinction between methods is whether the amount of time a respondent spends per question is recorded (and whether you can tell if they look up information!). If you are doing a survey on beliefs about outside options, for instance , it would be suspect if those with more accurate responses spent longer answering a question on average pay in their labor market. Some people who conduct internet surveys collect information on IP addresses and location (you generally have to specify this during the IRB process); this information can also be used to assess the quality of responses.

Reducing Non-Response

Many people do not respond to surveys. People who do not respond to surveys may differ from those who do in ways that can bias your results (think about recent political polls). An old study by Porst and Briel (1995) suggests that there are three broad reasons people respond to surveys: (1) altruism, (2) egoism, and (3) interest in the survey itself. People may be motivated to respond to surveys out of altruism—for instance you might respond to a government survey because it is your "civic duty" or you might complete a researcher's survey because you broadly think that research is important. People may simply enjoy surveys. They may also have a particular interest or belief in the importance of the topic you advertise that you are studying.

Because we may worry that individuals select into responding to a survey, when conducting a survey you should do whatever possible to increase response rates (overall). You should also (when feasible) embed variation that will allow you to disentangle different types of non-response. Intuitively, some people may not respond because they did not see your initial messaging. Others may not respond because they made an active choice not to engage. By changing the cost/benefit calculation for people (and by changing the salience of



the messaging) you may be able to identify the "survey compliers" and examine whether there is selection into your survey along the dimensions of interest. Note that if you survey people multiple times, you need to worry both about the initial response rates and attrition.

Be careful when comparing the response rates across papers. Researchers who use panels may report the fraction of individuals who responded conditional on having responded in earlier waves. But these individuals have already been screened on willingness to participate in surveys!

Messaging and Survey Details The first step to improving response rates is writing good recruitment messaging. Make sure that this is short, clear, and informative about what the study does and why it is important. You should also specify the length of time the survey will take. Ideally this will be short; I would expect response rates to decrease with the time you state (which should be truthful!). If you are conducting the survey on your own, think about how you want to identify yourself, and whether the way you state your credentials could lead to selection into non-response (or lead participants to think you want a certain answer). If you are running a survey through a firm, think about whether individuals will feel comfortable providing certain information to their employer (and, in particular, to the group within the firm you say is conducting the survey).

It is not common to vary the recruitment messaging used in a survey (one reason: it would be somewhat strange, and potentially unethical, for you to tell different individuals the survey does different things).

Financial Incentives One of the most common approaches to dealing with non-response is to provide individuals with financial incentives. Of course, what a "reasonable" financial incentive is depends on your target population and their average hourly wage.

For a fixed budget, you can vary the size of the per-person incentive. You can vary whether the incentives are pre-paid, regardless of survey completion, or are paid based on some level of engagement with the survey. You can also vary whether individuals get the incentive for sure, or with some probability (a lottery). Of course, a lottery will give you more control over the final budget. However, individuals may (and should) perceive lotteries differently.

My read is that there is no clear guidance on what size of incentive maximizes response rates within a given population. There may be some evidence that the relationship between the amount of the incentive and non-response is non-linear. Some of the older literature suggests that financial incentives may be effective in getting lower-income individuals to participate. More recent papers caution against providing financial incentives (see, e.g., Dutz et al., 2021). Evidence is mixed on whether those induced to participate by financial incentives provide lower-quality answers. My general advice would be to randomize the financial incentives you provide across participants and to have multiple levels of financial incentives that you randomize. The variety of estimates in the literature suggests that the response to financial incentives may vary across populations, surveys, and incentives.

Endorsement Letters Another "free" approach to increase response rates is to try to improve response rates by changing individuals' beliefs about the credibility/importance of your research. This is sometimes done through the use of endorsement letters. For instance, Bloom and Van Reenen (2007) obtained endorsements from governments in a variety of countries. They provided this information in their initial solicitations in order to convince firms to participate. DHST did not provide endorsement letters. However, they conducted their survey through the IAB, a research organization within the Federal Employment Agency.

In CHH we sent endorsement letters to a random fraction of those included in our survey. The letters did not have any detectable impact on response rates but were a low-cost attempt to produce additional variation. (The cost was basically the printing cost + the time to write the letter and get it signed.) It is possible that the letters did not have a huge effect because we were already fielding the survey through a respected organization in Germany—the IAB.

Randomized Follow-Up Randomized follow-up reminders to non-respondents boost response rates and can allow you to distinguish between active and passive decisions to participate (see Dutz et al., 2021). Most people who run surveys will tell you that reminders are the most effective way to boost response rates.

IRB

You'll need to complete an IRB. It usually makes sense to do this early in the process once you've figured out the logistical details. You can submit revisions if you change the questionnaire. You will typically need to provide detailed information on sample sizes, informed consent, data security, and compliance with GDPR (if applicable).

2.4 Attrition and Data Quality

Getting people to start your survey is only the first step. Next, you want them to complete all of the questions and to do so *meaningfully*. You do not want people to simply skip through the survey or click things at random.

Do not underestimate the value of providing respondents with a smooth experience. If you are doing a web-based survey, have a pleasing visual interface for both computers and mobile devices, optimizing the amount of text that appears on the screen at one time. If you are doing a written survey, there are similar considerations. Also be aware that sensitive questions may lead individuals to leave your survey. You should try to identify whether any questions you are asking are likely to be sensitive or personal. People often put these questions later in a survey so that they can examine differential attrition based on these questions.

To assess the quality of responses, you can embed "attention checks" into your survey. You should trade off the value of attention checks with the cost of having a longer survey: I would recommend only including a couple. Attention checks may not be acceptable in all contexts; I probably would not include them in a survey of CEOs!

To ensure the quality of individual responses you can provide monetary incentives to answer questions correctly. This is not feasible if you cannot verify the accuracy of individuals' responses in real time, is not acceptable to all survey partners, and may prompt respondents to look up the correct answer. You should consider whether you will be able to check whether respondents used outside information to answer.

Of course, not everyone will respond to your survey. And those who do respond may skip questions or drop out at a certain point. You should provide information on both overall response rates and attrition throughout the survey. Ideally, you won't find any pattern of selection into your survey. If you want to selection-correct your estimates, however, you can use probability re-weighting or parametric approaches.

It is common practice to restrict your sample to "complete" responses (individuals who made it to at least a certain point in the survey). However, you should check that your results are robust to changing this definition. If you've followed the advice above, you also

have plenty of ways to examine whether incentive compliers differ in their responses!

3 Surveys: Questionnaire Design

3.1 General Advice

One of the more practical considerations when writing a survey is how to word the questions. Should you allow open-ended questions (text fill-in)? Should you allow individuals to choose multiple responses? How should you order the options? Should you include an "other"? If you are running an in-person or phone survey, what should your interviewers ask? How should they respond to questions?

If you start running your own surveys, you should invest some time in learning best practices. There are a number of books on survey design that I am happy to recommend. If you are running an interactive survey (where interviewers talk to interviewees), you will likely find the information on the World Management Survey website particularly useful. The authors provide all of the resources they use to train interviewers. You will see that this is fairly involved (for good reason).

Before I draft a final set of questions, I do 'pre-piloting' via (1) conversations with people who look like my potential respondents and (2) low-cost online pilots. You can do things like:

- Present open-ended versions of questions not intended to be open-ended. Check to make sure all of the most frequently mentioned items are covered by the planned options.
- Especially in 1:1 conversations, ask if the questions are sensitive in any way or are difficult to answer.
- Present people with scenarios and ask them how they would respond. Check that their answers line up with what you expect (if not, this reveals you may not be eliciting what you think you are eliciting).
- Ask people for feedback on the question wording (i.e., whether the wording is confusing and how it could be improved). Since many of the people on online platforms take a lot of surveys they sometimes do have feedback on this!

3.2 List for Non-Interactive Surveys

For non-interactive surveys, I **strongly** recommend that, before you field a pilot, you do the following two things:

1. Go through the questionnaire question by question and indicate how you would code up the variables based on the responses. If you know you will group certain responses, you may want to reduce the number of options given to respondents. This may mean converting open fields to multiple choice questions or reducing the number of choices. **Every** question should have a purpose and have exhaustive answers. Longer surveys have more attrition; more complicated question have higher non-response rates.

    (a) E.g., why would you ask people to choose between whether they got 0 offers, 1 offer, 2 offers, 3 offers, 4 offers, 5 offers, 6 offers, ... if all you care about is whether they got 0 offers, 1, offer, or multiple offers (and if this version captures most of the relevant variation)?

    (b) If you are eliciting things that are "standard" in the literature, just use existing questions (with citation).

2. Check each question to ensure

    (a) The responses are "ordered" the same way (yes always comes before no, higher responses always mean the same thing?).

    (b) The answer choices are unique and exhaustive. Note that a general rule of thumb is not to have more than 4 or 5 potential responses. Use an other category if you need to.

    (c) None of the language is 'leading'

    (d) Each question asks one (and only one) thing.

    (e) Each question has one (and only one) interpretation, which lines up with what you want to ask.

    (f) There are no technical terms or jargon.

3. Consider whether respondents will consider each question to be sensitive

    (a) Would answering the question truthfully generate any risks?

    (b) Are the demographic questions so detailed that respondents are identifying themselves?

4. Consider both open-ended and closed-ended questions.

    (a) Open-ended questions can provide useful information, but are costly (and often have low response rates). Limit the number.

5. Consider question order effects: the way a respondent answers a question may be shaped by their answers to previous questions.

    (a) It's generally best to start with general questions that are easy to answer.

    (b) If you elicit multiple numerical responses in a row later numbers may be influenced by earlier numbers. It is difficult to separate survey anchoring from true anchoring in beliefs! In general, topics or numbers mentioned earlier in a survey can impact respondents' answers to later questions.

    (c) Randomizing the order of questions can help improve your data.

6. Consider whether (and how many) attention checks you need.

    (a) Attention checks may not be appropriate in some populations (e.g., if you are interviewing CEOs, you might not want to do this). In many (e.g., internet panels), they may be critical.

    (b) To keep the questionnaire as short as reasonably possible, don't have too many attention checks. Try to focus on a small number of highly effective checks.

    (c) Reduce the survey burden wherever possible (i.e., the amount of attention respondents need!). Keep a clean survey format and be consistent in how you order potential answers (i.e., highest to lowest, positive to negative).

7. Use filtering and branching

    (a) Try not ask respondents questions that do not apply to them. Use filters and branching.

8. Pre-test your questionnaire, to determine:

    (a) Whether you are missing any cases in your multiple-choice questions. As I mention above, you can ask individuals open-ended versions of your multiple-choice questions and verify that the responses map onto your cases.

    (b) How respondents view potentially difficult questions. Are people likely to skip the question or just choose at random? How long do they typically spend on that question page (if online) or on answering the interviewer?

(c) How do respondents interpret potentially ambiguous questions. Are there multiple interpretations?

9. Run an official pilot and check

   (a) When people drop out of the survey: Are there particular questions that seem to trigger drop out? Can these be eliminated or reworded?

   (b) Breakdown of responses for multiple-choice questions: Is there variation in the responses? If you have different buckets for the outcomes, are you isolating variation in the relevant range?

   (c) How missing values are recorded by your survey software or by your interviewers.

   (d) How the results of any in-survey randomizations are recorded by your survey software or by your interviewers.

   (e) The response rates for particular questions: do people often leave questions of interest blank?

   (f) If you ask related questions, do the answers line up?

10. Specific notes for internet surveys

    (a) Radio buttons are better than drop-down menus

    (b) Randomizing question order is good (easier in some software than others). While you can randomize the order of responses (e.g. high–>low, vs low–>high) do not do so within respondent, only across respondents.

You should also, of course, do the survey yourself many times (in addition to generating fake data through the survey software)—going through all of the options to make sure the survey flows as it should. One useful thing to do is generate fake data yourself by taking the survey multiple times—then try to write code to clean the responses. Are the data recorded as you expect? How does the software you are using code missing values? If you write code to clean the responses now, you will have it ready when the survey closes!

Something will inevitably go wrong in your pilot. You may realize some questions are not as useful as intended, or that there are logistical issues in the launch. You will therefore almost certainly make some minor changes between the pilot and the full launch. I strongly recommend that you build in time for a "soft launch" before fielding the full version. In other words, do not send the survey to everyone at once; send it to a small, random subset of your final respondents just to clarify that there are no issues. As an example, if you are mailing a bunch of survey invitations, send a small batch out first.

3.3 Eliciting Valuations

It would be impossible to cover all of the details of question design in these notes. Here I go over methods for estimating WTP for things (a thing economists often want to do!). The goal is both to teach you best practices for this use case, and to suggest that, in many cases, there are different ways of eliciting the same thing. Some approaches generate higher quality data than others. You should invest some time learning whether there are best practices for the types of questions you want to field. Economists undervalue the importance of writing good questions!

There are three general approaches to identifying individuals' willingness to pay:

1. Contingent/stated valuation (CV approach)
2. Discrete choice experiments (DCEs), sometimes called the 'take it or leave it' approach
3. Becker-Degroot-Marshak (BDM) mechanism

The CV approach is perhaps the most intuitive: you just ask people what they would be willing to pay (potentially under a certain set of conditions). In theory this gives you a point estimate of the WTP for each respondent. In discrete choice experiments you offer participants the opportunity to buy the product at a randomized price, and then using the observed data to estimate mean willingness to pay. This bounds the willingness to pay for a given individual; you use the full sample to estimate average valuations. To implement BDM you ask participants the maximum they are willing to pay, and then randomly generate a price for them. If the randomly generated price is not more than their maximum willingness to pay, they buy the object for the generated price; otherwise, they do not buy the object. This mechanism is essentially a second-price auction, with the usual highest bid of the other players replaced by the randomized price.

The labor literature has focused on discrete choice experiments as the standard for eliciting valuations. All recent 'top' papers on compensating differentials use this approach (see Mas, 2025, for an overview of this literature). The set-up is easy for participants to understand, respondents may not know their precise valuation: if I asked each of you the outside salary at which you would take a data scientist offer from {Specific Firm}, would you know the exact number? Or do you just have a general sense over the ranges? While the CV approach makes it 'seem' like you have an explicit valuation for each individual (rather than the bound on the valuation the DCE generates), it is not clear this additional precision is reliable. A recent paper suggests the bias of the CV approach can be on the order of 1000% (Rodemeier, Forthcoming).

People sometimes refer to DCEs as stated choice experiments because they rely on respondents' stated (rather than revealed) preferences. Of course, stated choice experiments are still a second best to revealed choice experiments. But we can't always run the experiment we want to run. There is evidence that, at least in labor, DCEs can yield valuations which are similar to revealed choice experiments (Mas and Pallais, 2017).

Under some assumptions, BDM can (like the CV method) tell you the exact willingness to pay (Dupas and Miguel, 2017). However, there is evidence that BDM data may be less accurate than the DCE approach (though potentially more accurate than the CV approach) (Berry, Fischer and Guiteras, 2020). This could be driven by the fact that BDM is more complex for respondents (who might not understand it is a dominant strategy to bid their own valuation). Unless you really care about individual-specific valuations, I might just use a DCE; this is what most people do in the literature (see, e.g., Mas and Pallais, 2017; Wiswall and Zafar, 2018).

4 Surveys: Validation

4.1 Piloting and Pre-Testing

When developing your questionnaire you should engage in substantial piloting and pre-testing. This is particularly true if you are asking "difficult" questions or questions that are substantively different from what other researchers have asked.

For instance, if you are surveying human resource managers you may find it useful to engage in in-depth interviews with people from that population. You may want to also have them go over your survey and explain to you how they interpret various answers. You may ask them open-ended versions of the questions you want to ask—and ensure that the full set of responses can be mapped on to the options provided in your questionnaire. You can also pose hypothetical scenarios (where you are sure of how you would want them to respond) and see whether their responses map on to the provided answers in the way you expect.

This is not a substitute for piloting. You will also want to conduct a formal pilot.

4.2 Ex Post Internal and External Validations

You should validate your survey, even if you are not trying to produce a "new" measure (as Bloom and Van Reenen (2007) were).

You can conduct "internal validation" tests by comparing within-respondent answers to different questions. For instance, if you elicit similar information in different ways you may correlate the within-person responses. You can also re-interview people or interview other

people who should have the same information. Bloom and Van Reenen (2007) collected a second set of responses (different plant manager at a different location) for 64 of the firms in their sample. They then showed that there was a strong correlation in responses, inconsistent with a pure measurement error story.[1] In Caldwell, Haegele and Heining (2025) we validated a summary measure of workers' outside options by showing it was correlated with more objective measures (number of job offers received, etc.).

You can conduct "external validation tests" by comparing things in your survey to known things about the respondents. For instance, Bloom and Van Reenen (2007) showed that their management scores correlated with various firm-level measures of performance. DHST show that time spent on job finding does change in predictable ways when individuals get jobs (as measured in the administrative IAB records). If you collect information that others have asked—e.g., risk aversion—you can compare the distribution in your sample (and, for instance, gender differences) to what people have found in similar populations. This is yet another reason to follow precedent when measuring things that are common in the literature: it enables ex post comparisons.

[1] They faced a trade-off between the size of the overall sample and the number of firms for which they collected multiple responses. Having more firms allowed them to examine heterogeneity in their estimates (remember standard errors fall proportionally to $\sqrt{n}$, not n). If measurement error is not a big issue, this is likely a better use of money.

References

Berry, James, Greg Fischer, and Raymond Guiteras. 2020. "Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana." *Journal of Political Economy*, 128(4): 1436–1473.

Bewley, Truman F. 1999. *Why Wages Don't Fall During a Recession.* Harvard University Press.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *The quarterly journal of Economics*, 122(4): 1351–1408.

Caldwell, Sydnee, Ingrid Haegele, and Joerg Heining. 2025. "Bargaining and Inequality in the Labor Market."

Cullen, Zoe B, Shengwu Li, and Ricardo Perez-Truglia. 2022. "What's My Employee Worth? The Effects of Salary Benchmarking." National Bureau of Economic Research.

DellaVigna, Stefano, Jörg Heining, Johannes F Schmieder, and Simon Trenkle. 2022. "Evidence on job search models from a survey of unemployed workers in germany." *The Quarterly Journal of Economics*, 137(2): 1181–1232.

Dupas, Pascaline, and Edward Miguel. 2017. "Impacts and determinants of health levels in low-income countries." In *Handbook of economic field experiments*. Vol. 2, 3–93. Elsevier.

Dutz, Deniz, Ingrid Huitfeldt, Santiago Lacouture, Magne Mogstad, Alexander Torgovitsky, and Winnie van Dijk. 2021. "Selection in Surveys." National Bureau of Economic Research.

Handel, Benjamin R, and Jonathan T Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." 105(8): 2449–2500.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer. 2024. "Biased Beliefs about Outside Options." *Quarterly Journal of Economics*.

Mas, Alexandre. 2025. "Non-Wage Amenities." , ed. Christian Dustmann and Thomas Lemieux Vol. 5 of *Handbook of Labor Economics*. Amsterdam, NL:Elsevier.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review*, 107(12): 3722–59.

Porst, Rolf, and Christa von Briel. 1995. "Wären Sie vielleicht bereit, sich gegebenenfalls noch einmal befragen zu lassen? Oder: Gründe für die Teilnahme an Panelbefragungen." *ZUMA-Arbeitsbericht*, 95(4).

Reynolds, Lloyd. 1951. *The Structure of Labor Markets.* Harper and Brothers.

Rodemeier, Matthias. Forthcoming. "Willingness to Pay for Carbon Mitigation: Field Evidence from the Market for Carbon Offsets." *Review of Financial Studies.*

Stantcheva, Stefanie. 2022. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible." National Bureau of Economic Research.

Wiswall, Matthew, and Basit Zafar. 2018. "Preference for the Workplace, Investment in Human Capital, and Gender." *The Quarterly Journal of Economics*, 133(1): 457–507.