

**Working in the Shadow:
Survey Techniques for
Measuring and Explaining
Undeclared Work**

Lilith Burgstaller, Lars P. Feld, Katharina Pfeil

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Working in the Shadow: Survey Techniques for Measuring and Explaining Undeclared Work

Abstract

Little is known about the size and determinants of undeclared work. While approaches to measure the shadow economy have been extensively discussed, conventional surveys dominate research on undeclared work. We review and extend this literature by first referring to the most recent survey data on undeclared work in Germany and, second, by discussing four experimental survey techniques as well as their few applications to questions of undeclared work. We argue that the randomized response technique and list experiments would validate and improve prevalence estimates of undeclared work, whereas careful design of information provision experiments and discrete choice experiments would fill the gap on determinants that causally affect decisions to supply and demand undeclared work.

JEL-Codes: H260, E260, O170, D910.

Keywords: undeclared work, experimental survey, survey data.

Lilith Burgstaller
Walter Eucken Institute
Goethestraße 10
Germany – 79100 Freiburg
burgstaller@eucken.de

Lars P. Feld
Walter Eucken Institute
Goethestraße 10
Germany 79100 Freiburg
feld@eucken.de

Katharina Pfeil
Walter Eucken Institute
Goethestraße 10
Germany – 79100 Freiburg
pfeil@eucken.de

May 11, 2022

1 Introduction

Quantifying the magnitude and determinants of undeclared work is methodologically challenging, but highly relevant for designing policy instruments that provide incentives for declaration and deter under-reporting. Undeclared work is widespread in household-related services, for example in cleaning, child care and elderly care (OECD, 2021). Also in construction, hospitality and retail or repair services, work without reporting taxes or social security contributions is common (European Commission, 2020). Understanding the nature of undeclared work is important, as it leads to losses in tax and social security revenues, pensions, unemployment benefits, redundancy payments, unpaid wages and unfair competition (Feld & Larsen, 2012b). However, measuring the prevalence of undeclared work is difficult because by definition it takes place in the shadow. This paper reviews conventional survey evidence of undeclared work as well as experimental survey techniques that may increase the validity of the measurement of undeclared work and allow for the causal identification of its determinants.

Undeclared work is defined as a taxable and essentially legal economic activity that is compensated with income, but that is deliberately not reported to the tax, social security or customs authorities. Thus, taxes, social security contributions and other fees are not deducted, and labor market standards such as minimum wage regulations or administrative obligations are avoided.¹ Undeclared work and tax evasion have some overlaps, but are not identical. The latter also contains individual under-reporting of capital income tax or transfer income. In turn, the focus of undeclared work lies on the shared awareness and agreement of buyers and sellers to benefit by concealing the activity from the authorities (Feld & Larsen, 2012a). This means that we can also understand many (but not all) cases of undeclared work as “collaborative” or “collusive” tax evasion.

There are macro and micro methods to estimate the extent of the shadow economy or undeclared work in particular.² Macro approaches include the monetary transaction method, the currency demand approach, the income gap method, the electricity approach, the discrepancy method and the MIMIC estimation procedure (see Dybka et al., 2019; Giles et al., 2002, for recent applications). Micro approaches include surveys with direct or indirect questioning, estimations with tax audit data as well as laboratory and field experiments (see Žukauskas & Schneider, 2016; Putnins & Sauka, 2021; Doerr & Necker, 2021, for recent applications).

¹Undeclared work only covers a part of the shadow economy because it does not include illegal activities (see Schneider & Enste, 2000; Feld & Schneider, 2010).

²The literature often distinguishes between direct and indirect methods, where indirect methods can use micro or macro data. However, as there are also indirect questioning techniques in surveys, e.g., asking about tax morale or undeclared work in the peer group, we decided to differentiate micro and macro approaches. More detailed overviews of the methods, their advantages and disadvantages can be found in Schneider (2021); Slemrod & Weber (2012); Schneider & Enste (2000); Schneider & Buehn (2018); Feld & Larsen (2012b); Kirchgässner (2017); Feld & Schneider (2017).

None of these approaches come without weaknesses. Macro methods tend to overestimate the size of the shadow economy, whereas micro approaches provide lower-bound estimates (see Schneider, 2021; Schneider & Buehn, 2018). Macro approaches are sensitive to their underlying assumptions and calculations. They face the “double counting problem”, because they also include legally-bought material, and, as in the case of MIMIC, only provide relative coefficients instead of absolute values, relying heavily on the calibration procedure (e.g., Schneider & Buehn, 2018). Importantly, they present overall trends of the shadow economy, but cannot differentiate between the shadow economy and undeclared work.

The survey technique is particularly suitable to delineate undeclared work from other activities within the shadow economy (Schneider, 2021).³ Recent studies have used survey-based methods as well as survey data complemented with administrative data to estimate the fraction of undeclared work (Williams et al., 2017; Elek & Köllő, 2019; De Gregorio & Giordano, 2016). However, survey data may suffer from dishonest answering behavior, referred to as social desirability bias. This means that answering behavior can be distorted by self-presentation concerns, making respondents omit or lie about socially undesirable decisions (Krumpal, 2013). Besides biased answering behavior, non-responses limit the measurement of the true magnitude of undeclared work in surveys. Moreover, the estimates may depend on the exact wording of the questions and the sample included in the survey, e.g., households or company managers. Nevertheless, surveys have the advantage to collect data on the structure and determinants of undeclared work that other methods cannot measure (Feld & Larsen, 2012a).

In addition to surveys, questions of tax compliance are studied in laboratory and field experiments.⁴ It is important to emphasize that the experimental literature has long focused on individual tax evasion and has only lately begun examining collaborative aspects of evasion.⁵ Lab experiments allow researchers to cleanly disentangle the determinants of tax evasion as they manipulate the variables of interest exogenously (e.g., audit rates or fines). Even though laboratory experiments provide high internal validity, drawbacks are the lack of external validity and the fact that participants are often recruited from a student population. Due to these weaknesses, natural field experiments have been increasingly used in recent years (see Hallsworth, 2014; Slemrod & Weber, 2012). The advantage of natural field experiments is that subjects do not know about the experiment. Researchers observe how participants respond to the incentives presented to them in their natural environment (Harrison & List, 2004). This minimizes

³Surveys can further distinguish between undeclared payments that are in kind or in cash.

⁴See Slemrod & Weber (2012) for an overview and Alm & Malézieux (2021); Antinyan & Asatryan (2020) for recent meta-studies on tax evasion experiments in the laboratory and field.

⁵Recently, the literature on determinants of joint tax evasion in field experiments (Doerr & Necker, 2021; Bjørneby et al., 2021) and lab experiments (Burgstaller & Pfeil, 2022; Lohse & Simon, 2021; Fochmann et al., 2021; Balafoutas et al., 2015; Abraham et al., 2017) has been growing immensely.

biases that may result from individuals feeling observed and responding with socially desirable behavior. However, because field experiments are implemented in a specific time and setting, their results have limited generalizability and the magnitude of undeclared work for a whole economy cannot be estimated. As all approaches have their strengths and weaknesses, the choice of one approach over the other is ultimately dependent on the research question.

In this paper, we focus on survey techniques because they provide individual data on undeclared work and include determinants such as perceived sanctions, audits and social norms. In the first part, we provide a comprehensive overview of the development of undeclared work in Germany. We summarize representative survey data, covering the years from 2001 to 2019. The case of Germany serves to demonstrate the strengths and weaknesses of conventional surveys. Next, we argue that while previous surveys are a good starting point for presenting the factors influencing undeclared work and its structure, this technique has room for improvement. First, we do not know to what extent social desirability bias may distort the estimates. To date, there have only been few attempts at quantifying the potential bias regarding questions of undeclared work (Kirchner et al., 2013; Trappmann et al., 2014, are the exceptions). Second, causal evidence on the determinants of undeclared work is lacking. Hence, in the second part of the paper, we present and discuss several experimental survey techniques to improve survey designs. While the application of such survey designs is beyond the scope of this paper, we outline possibilities of how to apply the proposed methods to studying undeclared work and suggest research questions that should be addressed in the future.

This paper contributes to our understanding of the variety of experimental methods to examine undeclared work. Even though a discussion of selected experimental survey techniques exists in the literature (John et al., 2018; Wolter & Diekmann, 2021; Höglinger & Jann, 2018; Trappmann et al., 2014; Kirchner et al., 2013; Krumpal et al., 2018), a thorough review and comparison of such techniques and their applicability to questions of undeclared work has been missing so far. We argue that experimental survey techniques may, on the one hand, lead to an improved measurement of undeclared work and, on the other hand, allow for causal identification of its determinants. By emphasizing the potential of experimental survey techniques, we aim to motivate more rigorous research in this area that moves beyond the mere measurement towards an explanation of why undeclared work is prevalent.

This paper proceeds as follows. In Section 2, we will summarize and compare existing survey evidence on undeclared work. We take stock of the participation rates and the potential of undeclared work in Germany and report data on perceived deterrence as well as social norms. Section 3 gives an overview of experimental survey techniques such as the randomized response techniques, list experiments, information provision experiments and discrete choice experiments and their applications to questions of (collaborative) tax evasion and undeclared work. Section 4 concludes.

2 Survey Evidence on Undeclared Work in Germany

In this section, we give an overview of the development of undeclared work in Germany, measured by the conventional survey method. Most studies use survey data from the Special Eurobarometers on undeclared work, which were conducted in 2007, 2013 and 2019.⁶ Feld & Larsen (2012a) are among the few who have examined undeclared work in individual, representative surveys in Germany. We use their data and focus on the case of Germany because it provides comprehensive time-series data for the years 2001 to 2012 (N=20,059). To extend the period of observation, we further include the latest Eurobarometer data from 2013 and 2019. All data was collected in face-to-face computer assisted personal interviews.

The survey flow is designed such that more general, non-sensitive questions about do-it-yourself work or household-related services lead into sensitive questions about undeclared work. The transition between these two parts includes a description of undeclared work to ensure that respondents understand it.⁷ Moreover, interviewers are instructed to emphasize confidentiality and to mention the scientific purpose of the survey. After this introduction, respondents are asked to indicate whether they supplied undeclared work during the previous twelve months. Questions on the hours worked, the hourly wage, the general willingness for offering undeclared work as well as perceptions about deterrence measures and social norms follow. The survey ends with various socio-demographic questions.

In Table 1, we report the main variables for the supply of undeclared work in Germany over all survey waves. Participation is a measure of the prevalence of undeclared work and denotes the percentage of respondents that have engaged in undeclared work in the previous year. We find that the participation rate has strongly decreased since 2007. While respondents report participation rates around 10 % from 2001 until 2008, the reports are only at 3.5 % in 2012. The two Eurobarometer waves in 2013 and 2019 confirm this trend with 2.8 % in 2019. 2006 can be considered an outlier with a significantly lower reported fraction of undeclared work as compared to other years (Feld & Larsen, 2012a). This downward trend in the supply of undeclared work does not only coincide with the introduction of a tax subsidy for private households, but also with the

⁶Current papers that use Eurobarometer data are Williams & Öz-Yalaman (2021); Horodnic & Williams (2022); Williams & Nadin (2014). For papers using self-collected survey data, see Žukauskas & Schneider (2016); Putnins & Sauka (2021).

⁷Feld & Larsen (2012a, p. 86) display the following definition: “The next questions are about what is popularly called ‘undeclared work’. There is considerable evidence to show that a large part of the population accepts ‘undeclared work’ and ‘undeclared trade in goods’ – i.e. activities which circumvent the Inland Revenue, where all parties benefit because they do not pay tax or VAT, etc. This can involve ‘undeclared activities’ which you pay for in cash, but can also include reciprocal favors between friends, acquaintances and family members.”

positive development of the overall economy in Germany and increased deterrence measures (Feld et al., 2013). At the same time, the general willingness to offer undeclared work, i.e. the potential, remains at the same level of around 18 % of respondents stating to be willing to work without declaration if they have the chance.

For average hours of work without declaration, we also observe a decrease, from 8.2 hours in 2001 to 5.8 hours in 2012. The hourly wage for undeclared work is fairly stable around 11 Euros on average. Reported wages increase to 13.3 Euros in 2005, but drop again to 11.1 Euros in 2012.

Table 1: Overview of the supply of undeclared work in Germany 2001-2019.

	Participation	Potential	Hours	Hourly Wage (€)	N
2001	10.4	17.3	8.2	10.3	5.686
2004	8.8	17.1	7.5	10.4	2.143
2005	11.1	19.8	6.7	13.3	2.144
2006	5.8	15.5	6.4	11.8	2.176
2007	9.7	18.2	6.3	12.8	1.096
2008	9.6	–	5.1	–	2.122
2009	6.0	20.9	–	–	2.149
2012	3.5	19.8	5.8	11.1	2.543
2013	2.1	–	–	–	1.499
2019	2.8	–	–	–	1.565

Notes: The six survey waves 2001 to 2008 were first reported in Feld & Larsen (2012a). The data for 2009 is reported in Feld et al. (2010) and for 2012 in Feld et al. (2013). Data for 2013 and 2019 is retrieved from Eurobarometer 402 and 498, see European Commission (2014, 2020). Participation and potential are displayed as percentages. Minutes are expressed as percentages of an hour.

Questions: “Have you carried out activities of this kind during the past year?” (*Participation*), “If you had the opportunity, would you engage in undeclared work?” (*Potential*), “If yes, how many hours did you work on average per week?” (*Hours*), “If yes, how many euros did you receive on average per hour?” (*Hourly Wage*).

In addition to the prevalence of undeclared work, the survey data allows us to examine various factors that affect the decision to work without declaration. We can categorize aspects that are enquired in surveys into deterrence measures and social norms. The former refers to perceptions of the detection probability and of the severity of fines, whereas the latter is concerned with the acceptance of various kinds of undeclared work. Regarding deterrence, Table 2 shows the perceived risk of being detected if carrying out undeclared work. A large fraction of around 40 % of respondents assesses the risk

Table 2: Perceived risk of detection for undeclared work, in %.

	Very high	Fairly high	Fairly small	Very small	Don't know/ refusal
2001	10.8	25.4	30.1	18.2	15.6
2004	9.4	29.7	32.3	15.1	13.5
2005	11.5	31.3	32.4	13.8	11.1
2006	11.7	30.3	30.9	16.3	10.9
2007	12.0	28.7	34.8	12.6	12.0
2008	12.1	25.1	33.9	17.1	11.8
2009	11.5	36.8	40.4	11.3	-
2012	13.1	33.9	32.5	11.4	9.7
2013	5.3	28.9	41.7	14.0	10.5
2019	6.4	28.8	48.4	8.6	7.8

Notes: The six survey waves 2001 to 2008 were first reported in Feld & Larsen (2012a). The data for 2009 is reported in Feld et al. (2010) and for 2012 in Feld et al. (2013). Data for 2013 and 2019 is retrieved from Eurobarometer 402 and 498, see European Commission (2014, 2020). *Question:* “People who carry out undeclared work risk the authorities finding out and issuing supplementary tax bills and perhaps fines, etc. In your opinion, what level of risk do people who carry out undeclared work run? Is it very high, fairly high, fairly small, or very small?”

of detection as high (very high or fairly high). It is important to note, however, that in the Eurobarometer surveys in 2013 and 2019, the risk of detection is perceived as comparably smaller. This could be due to the survey design, a change in politics or a change in perception over time.

The second aspect of deterrence is the perception of fines that can vary with the amount of undeclared work (see Table 3). In the surveys, different types of sanctions are categorized into payment of “taxes due”, payment of “taxes due and fine”, “imprisonment” and “all three”. Expected sanctions increase over time and with the generated monetary value of undeclared work. While in 2004 only 19.9 % expect that payment of taxes due are the sanction for undeclared work of 2,500 Euros, this fraction increases and is at 28.1 % in 2012. For undeclared work to the value of 50,000 Euros or 250,000 Euros, the fraction of respondents who expect imprisonment or all three types of sanctions increased slightly between 2009 and 2012. Moreover, the fraction of non-responses/refusals are much higher between 2004 and 2008 with around 25 % than in 2012 with around 11 %, suggesting that respondents have developed clearer expectations about sanctions for undeclared work over time.

To enquire social norms around undeclared work, respondents were asked to indicate to what extent they found certain types of behavior (un-)acceptable (see Table 4). Among the examples of behavior were welfare fraud, fare-dodging on the bus or train, different types of undeclared work and tax evasion. Private demand for private undeclared work is assessed as unacceptable by the smallest fraction of respondents (20.3 -

Table 3: Expected sanction for undeclared work, in %.

Year	Taxes due	Taxes due and fine	Imprisonment	All	Don't know/ refusal
<i>... to the value of 2,500 Euros</i>					
2004	19.9	51.2	4.9	-	24.0
2005	26.3	53.5	7.8	-	12.4
2006	28.1	44.1	-	-	27.8
2007	24.2	50.7	-	-	25.2
2008	29.8	45.2	-	-	25.0
2009	31.1	65.6	1.9	1.5	-
2012	28.1	58.1	2.0	0.7	11.0
<i>... to the value of 10,000 Euros</i>					
2006	10.5	48.7	11.6	-	29.2
2007	10.4	54.5	10.7	-	24.5
2008	9.5	53.6	10.0	-	26.9
<i>... to the value of 50,000 Euros</i>					
2009	6.8	74.7	11.9	6.6	-
2012	3.5	64.5	14.0	7.4	10.7
<i>... to the value of 250,000 Euros</i>					
2009	4.7	44.1	22.6	28.7	-
2012	1.8	31.8	22.5	32.8	11.2
2013	11.9	67.6	6.0	-	2.5
2019	11.1	72.7	7.5	-	0.6

Notes: The six survey waves 2001 to 2008 were first reported in Feld & Larsen (2012a). The data for 2009 is reported in Feld et al. (2010) and for 2012 in Feld et al. (2013). Eurobarometer 402 and 498 include general questions on the perceived sanctions without the category “all three”.

Question: “If the authorities find out that somebody has carried out undeclared work, the taxes due must be paid and perhaps the person will also become liable to fine or imprisonment. In your opinion, what sanction is to be expected if the authorities find out that somebody has carried out undeclared work to the value of 2,500 (10,000/ 50,000/ 250,000) Euros?”

27.6 %) as compared to the other types, whereas undeclared professional demand and supply are deemed the most unacceptable, similar to welfare fraud. Individual assessments of social norms for undeclared work are fairly stable until 2008, but in 2009 unacceptability increases. Moreover, in the Eurobarometer waves in 2013 and 2019, all types of behavior are considered more unacceptable. Private supply for private demand of undeclared work is also the most acceptable over all categories in the Eurobarometer, while professional demand and supply is the least acceptable.

It must be noted that the evidence presented in this section may serve to take stock of undeclared work in Germany, but it comes with a number of limitations. First,

Table 4: Individual assessment of various kinds of cheating behavior, absolutely unacceptable, in %.

	2004	2005	2006	2007	2008	2009	2013	2019
Welfare w/o entitlement	73.4	71.9	74.8	74.5	73.1	72.5	66.2	-
Free riding on bus/ train	51.4	48.4	51.8	50.0	49.4	-	45.6	-
<i>Undeclared work:</i>								
Carrying out	31.7	27.8	33.8	30.8	28.2	32.6	-	-
Private for private	24.7	20.3	26.5	23.9	22.0	27.6	30.9	46.7
Firm for private	47.9	41.8	45.9	43.8	42.6	50.5	58.8	70.6
Private for firm	62.8	62.0	63.8	59.7	59.7	64.3	67.9	76.5
Firm for firm	70.3	67.5	67.0	65.8	66.7	72.2	72.6	80.3
Tax evasion	45.8	44.1	50.5	46.4	47.2	48.5	53.3	66.5

Notes: The six survey waves 2001 to 2008 were first reported in Feld & Larsen (2012a). The data for 2009 is reported in Feld et al. (2010). Data for 2012 was not included due to limited comparability. Data for 2013 and 2019 is retrieved from Eurobarometer 402 and 498, see European Commission (2014, 2020).

Question: “Now I would like to know how you assess various behaviors. For each of them, please tell me to what extent you find it acceptable or not. Please use the following scale: ‘1’ means that you find it ‘absolutely unacceptable’ and ‘10’ means that you find it ‘absolutely acceptable’.”

surveys that aim to generate an estimate of the participation from a representative sample exclude individuals that do not have a permanent residence in the country of interest. This may bias the estimates downwards as they exclude certain groups not registered in Germany (see, e.g., Feld & Larsen, 2012a). Second, most of the estimates from existing surveys mainly capture the supply side of undeclared work. At the same time, the demand of undeclared work is under-investigated.⁸ Since undeclared work is a collaborative endeavour, the demand side is of crucial importance not only when analyzing the prevalence, but especially when identifying determinants of undeclared work or developing measures to combat it.⁹ Third, while each estimate gives insights as to the current state of undeclared labor in a given year, the comparability of different survey waves or survey instruments is limited. Slightly different phrasing or order of questions and varying definitions of undeclared work make comparisons between years

⁸Some exceptions are Williams & Kosta (2021); Williams & Martinez-Perez (2014); Horodnic et al. (2021) and Feld et al. (2013), who also present data on the demand of undeclared work.

⁹A challenge when examining both market sides is how to interpret potential discrepancies between supply and demand estimates of undeclared work (Feld et al., 2013). They may be ascribed to individuals without a permanent residence in Germany that are hard to capture in supply side surveys, to the fact that one supplier could work for multiple consumers or to individuals being more hesitant to reveal that they carry out undeclared work in comparison to demanding undeclared work.

more difficult.

These three shortcomings may be remedied by collecting data of suppliers and consumers in undeclared work on a yearly basis. The same survey instruments should be applied over a longer time span to assure comparability between survey waves. In addition to using a sample that is representative for the population, researchers should consider including groups of individuals that are not covered by representative samples but likely participate in undeclared work. Overall, data sets of such types may provide insight on the development of the prevalence of undeclared work over economic cycles and policy changes. Moreover, they may be used to exploit natural experiments that occur through policy changes over time or regional variation.

While these remedies may help to increase the validity of estimates of undeclared work, there are two further aspects limiting the explanatory power of conventional survey data. First, direct questioning in surveys likely generates lower bound estimates of undeclared work (Feld & Larsen, 2012a) as it relies on individuals admitting to their participation in concealed behavior. Comparisons of methods show that direct questions on sensitive topics tend to underestimate their true extent (e.g., Fisher, 1993; Tourangeau & Yan, 2007). This is a result of social desirability bias (Krumpal, 2013). As individuals aim to present themselves in the best possible way, they might not reveal their true behavior when it comes to stealing, cheating or undeclared work. Even more, participants may fear to be exposed for their law-breaking activities. When the participation in sensitive activities is enquired in direct interviews, social desirability bias may be even amplified by individuals wanting to please the interviewer (interviewer effects).¹⁰

Second and importantly, a causal estimation of the determinants of undeclared work is mostly not possible with conventional survey data. When it comes to the effect of sanctions, the risk of detection or social norms on undeclared work, survey data gives insights on correlations between variables. Even when conducted on a regular basis with a greater variety of samples and consistent questions, only natural experiments provide the necessary exogenous variation to examine causal effects. Both of these crucial limitations may be remedied by using experimental survey techniques.

¹⁰Feld & Larsen (2012a,b) report that results from Denmark point to interviewer effects meaning that interviewers' attitudes and expectations may have impacted responses regarding undeclared work.

3 Experimental Survey Techniques and Their Applications to Undeclared Work

3.1 Overview of Experimental Survey Techniques

Experimental survey techniques aim to provide unbiased estimates of sensitive behavior and their determinants. The randomized response technique and the list experiment may be applied to derive valid prevalence estimates of sensitive behavior and compare these with existing evidence from conventional survey data. Information provision experiments and discrete choice experiments may assess determinants of undeclared work causally, which may inform designing effective policy measures. The aim of this section is twofold. We first review four experimental survey techniques, show methodological advancements and discuss challenges and opportunities. Then, we survey the literature that applies experimental survey techniques to questions of undeclared work.

Randomized Response Technique. Research that aims at studying the occurrence of sensitive behavior in a population has been concerned with the question of how to generate valid estimates for a long time. A prominent method to derive truthful answers from participants is the randomized response technique (RRT) (Warner, 1965). This approach and its extensions have been used to study the occurrence of topics such as antisemitism in Germany (Krumpal, 2012), illicit drug use among athletes (Striegel et al., 2010) or corruption (Gingerich, 2010). The common idea to all variants of the RRT is to measure sensitive behavior on the aggregate level. Individuals' answers to a question cannot be directly linked to the sensitive question, which ensures individuals' anonymity and decreases subjects' reluctance to give truthful answers (Lee, 1993).

A conventional RRT is embedded in a structured interview. Participants receive a randomization device such as a coin. In the simplest version of the RRT, the forced response design, participants are instructed that a coin toss determines whether they answer the sensitive question or simply report "Yes" irrespective of the questions they have been asked. As the interviewer is unaware of the result of the coin toss, he or she does not know whether the interviewee answers "Yes" to the sensitive question or because of the result of the coin toss. Since the probability distribution of the coin toss is known, researchers may estimate the occurrence of the sensitive behavior from this exercise (Fox & Tracy, 1986; Gingerich, 2010).¹¹

While this approach grants anonymity to interviewees, it also has some drawbacks. Since estimates cannot be computed on the individual level but only on the aggregate, the statistical power of an RRT for a given sample is reduced in comparison to direct

¹¹In the above example, the prevalence rate of the sensitive behavior would be calculated by deducting 0.5 (the probability that the coin lands on a specific side) from the estimate of the RRT question and dividing this by 0.5 (Gingerich, 2010).

questioning. Moreover, insecurities about the design and aim of the questionnaire may distort participants' answers. If they are concerned that the interviewer thinks they are responding to the sensitive question, participants may be reluctant to answer "Yes" even if they are instructed to do so by the coin toss (Blair et al., 2015; Coutts & Jann, 2011). There have been multiple attempts to tackle the shortcomings of the original RRT (see Blair et al., 2015).

One variant of the RRT is the Crosswise Model (CM) (Yu et al., 2008; Korndörfer et al., 2014). As opposed to the traditional RRT, the CM does not require the presence of an interviewer but may be integrated in online or on-paper surveys. When using the CM, respondents receive a block of two questions containing one sensitive and one unrelated question. They are then asked to indicate whether their answers to these questions are the same or different. The probability distribution of the answers to the unrelated question should be known and the questions should not be correlated. This approach does not require participants to directly answer a sensitive question and therefore increases the validity of the estimate of sensitive behavior (Korndörfer et al., 2014).

When assessing the validity of the RRT and its variants, researchers have often relied on the "more-is-better-assumption" (Tourangeau & Yan, 2007). It implies that higher estimates of the behavior of interest depict the actual occurrence of the sensitive behavior better than lower estimates (Lensvelt-Mulders et al., 2005; Tourangeau & Yan, 2007). However, Höglinger & Jann (2018) challenge this assumption by referring to a considerable share of false positive reporting in surveys using the CM which leads to an upward distortion of the estimate.¹² Still, providing detailed explanations and implementing comprehension checks can reduce the rate of false positives as Meisters et al. (2020) report.¹³

List Experiment. Another experimental survey technique for dealing with sensitive questions is the list experiment (also: item count technique). The list experiment, first developed by Miller (1984), has received increased interest by researchers in the past for estimating the prevalence of sexual abuse (Traunmüller et al., 2019) or measuring attitudes against racial minorities (Kuklinski et al., 1997). Subjects receive a list of different behaviors and are asked to indicate a number denoting how many of these behaviors they have engaged in or are familiar with. The subject pool is randomly split into two groups. One randomly selected group receives a list with the sensitive item (long list), while the other group receives the same list of control items without the sensitive

¹²Höglinger & Jann (2018) evaluate the effectiveness of the RRT to reduce misreporting by comparing self-reported to actual cheating behavior in a dice game. While they find a decrease in false negatives (under-reporting) when using the CM as opposed to direct questioning, the fraction of false positives (over-reporting) increases.

¹³For a further discussion of the validity of the RRT see John et al. (2018) or Wolter & Diekmann (2021).

item (short list). If the two groups are homogeneous in the characteristics that affect the behavior in question, the prevalence of the sensitive item is the difference between the means of both groups (Droitcour et al., 2004).

When designing a list experiment, three to four non-sensitive items should be included in a list. Control attributes must be carefully selected to prevent ceiling and floor effects (Glynn, 2013). A ceiling effect occurs when participants are familiar with all the control items as well as the sensitive item in the list. For participants, indicating that they are familiar with all listed items is in effect the same as answering a direct question about the sensitive item. This could discourage them from responding truthfully. On the contrary, a floor effect occurs when participants have not experienced any of the control items but only the sensitive item. If they assume that this holds for other participants as well, they may be hesitant to indicate a “One”, as this would compromise their anonymity as well (Blair & Imai, 2012).

As in the case of the RRT, while deriving estimates on the aggregate level increases subjects’ anonymity, it also implies decreased statistical power for a given sample as compared to direct questioning. This shortcoming of list experiments may be remedied by applying the double list technique (Krumpal et al., 2018). In this variant of the list experiment, each subject is shown both, the long and the short list, in a randomized order. Hence, all subjects provide an answer to the sensitive item, doubling the effective sample size and allowing to complement individuals’ statements with their characteristics.

Similar to the RRT, efforts to validate list experiments compared to direct questioning need to rely on the “more-is-better-assumption”. While meta-analyses of list experiments generally find that they increase the number of reports of sensitive behavior, their validity depends on the study characteristics, such as research interest, items used or the mode of data collection (Li & Van den Noortgate, 2019; Ehler et al., 2021).

Information Provision Experiment. There is burgeoning literature using information provision experiments in various areas such as Corbacho et al. (2016) on corruption among households, Bansak et al. (2016) on attitudes towards refugees or Alesina et al. (2018) on the support for redistribution.¹⁴ An information provision experiment allows to study the effect of certain pieces of information or mental concepts on behavior and opinions (Haaland et al., 2022).¹⁵ Subjects are randomized into a treatment and a control group. In the treatment group, subjects’ attention is drawn to a (mental) concept by showing them information or asking them to remember a situation they have been in. Subjects in the control group receive neutral information or no information at all. Participants in both groups are then asked the sensitive question, such as whether they would engage in a certain behavior or how acceptable they find a given behavior.

¹⁴See Haaland et al. (2022) for an overview.

¹⁵In research that aims to prime subjects with psychological concepts instead of objective information, this approach is also called “priming” (Cohn & Maréchal, 2016).

The comparison of answers in both treatment groups allows to identify the effect of the primed concept (Cohn & Maréchal, 2016).

The sensitive question of interest is asked hypothetically in information provision experiments, which should reduce social desirability bias. However, to further reduce this bias, researchers may combine information provision experiments with list experiments or the RRT such that subjects do not have to state their (hypothetical) decision to participate in sensitive behavior. While hypothetical decisions likely decrease social desirability bias, they may also decrease the validity of results as participants' decisions do not entail consequences. This may be more severe when consequences entail financial losses or other punishment. However, Hainmueller et al. (2015) show that hypothetical decisions match real world behavior even when it comes to socially undesirable actions.

Moreover, when activating mental concepts, such as social norms, a potential pitfall is that researchers cannot be sure what they activated, i.e. they may measure the impact of a different concept than intended and cannot be certain what participants' reactions may be attributed to. In addition to that, hypothetical concepts such as an unknown policy instrument must be chosen in such a way that participants can easily understand them without being overwhelmed (Haaland et al., 2022). Using a variety of manipulation checks may help remedy these concerns regarding the design of the experiment (Cohn & Maréchal, 2016).

Discrete Choice Experiment. A discrete choice experiment (DCE) allows researchers to estimate individuals' utility from a certain product or service and the attributes that describe it (McFadden, 1986). They have been frequently applied in health economics (Poulos et al., 2011; Clark et al., 2014; De Bekker-Grob et al., 2010) or labor economics (Mas & Pallais, 2017; Datta, 2019; Humburg & Van der Velden, 2015). In a DCE, subjects have to decide between two or more options for a product or service in a choice set. Each option varies in several attributes that may take different levels. One participant usually faces multiple choice sets. This simulates actual decision-making behavior. In DCEs, the preference weight for each level of each attribute may be estimated. By including a monetary attribute (such as the price or wage) the willingness to pay (WTP) for each level of an attribute can be computed (Mas & Pallais, 2017). By incorporating the sensitive attribute of a service or product, researchers may examine whether the sensitive characteristic is an important determinant of their decision.

Horiuchi et al. (2021) emphasize that DCEs attenuate social desirability bias by two mechanisms. First, by including a sensitive item in a list of other items, subjects may not even be aware that they are being asked a sensitive question. Second, they may use the remaining attributes to justify their behavior. For example, if one of the options contains the potentially norm-violating attribute but seems superior in other attributes, subjects may rationalize their choice for that option.

Note that while the hypothetical character of these questions increases the probability

that a participant will respond truthfully, it also poses threats to the validity of the estimates obtained with DCEs and information provision experiments. When subjects' decisions are without consequences, they may decide differently than when their behavior entails monetary or other outcomes. However, as Mas & Pallais (2017) and Vossler et al. (2012) show, this problem may be mitigated by making the decisions in DCEs more consequential to participants. This may entail connecting subjects' experimental decisions to real world options, such as choices over two jobs in an actual recruitment process (see Mas & Pallais, 2017).

3.2 Applications of Survey Techniques to Study Undeclared Work

Experimental survey techniques can be powerful tools to advance our understanding of undeclared work. In the following, we illustrate the potential of such techniques by elaborating on the few existing applications to questions of undeclared work. Similar to our approach above, we structure the literature into applications into experimental survey techniques that estimate the prevalence of undeclared work on the one hand and studies on its determinants on the other hand. While there are few contributions that estimate the magnitude of undeclared work with either the randomized response technique or a list experiment, we found even fewer assessments of the determinants of undeclared work using information provision or discrete choice experiments.

Existing prevalence estimates are limited to list experiments on undeclared work among refugees in Germany (Doerr et al., 2022) and on the fraction of households acting as accomplices in VAT evasion in Columbia (Fergusson et al., 2019). Using an extension of list experiments, the item sum technique, Trappmann et al. (2014) studies the hours of undeclared work in Germany.¹⁶ To compare experimental survey techniques with direct questioning, Kirchner et al. (2013) conduct a study including the RRT and a list experiment with a German sample of individuals and firms to examine their experience with undeclared work. In another application of the RRT, Kundt et al. (2013) analyze the prevalence of under-reporting of sales and payment of envelope wages in Serbia. These estimates provide a fragmented picture of undeclared work in specific sectors and samples. Moreover, they use different questions which hampers the comparability and cross-validation of the estimates.

To the best of our knowledge, there are only a handful of papers implementing DCEs for a causal analysis of determinants of undeclared work. These are limited to evaluating the effectiveness of incentive schemes to reduce the demand of undeclared

¹⁶The item sum technique does not enquire whether different behaviors are carried out, but how much time participants spent on these behaviors. If the sensitive behavior is included only in the long list, such approaches allow to estimate the hours spent doing sensitive behavior, such as working without declaration.

work in private households. Adriaenssens et al. (2021) and Theys et al. (2020) focus on the Belgian service voucher policy and study households' preferences for declared work in a DCE. Burgstaller et al. (2022) combine a DCE with an information provision experiment and study the effectiveness of different tax credits in Germany.

Apart from this, information provision experiments have not yet been applied to questions of undeclared work, presenting a gap in the literature. Therefore, we draw on recent applications from the tax evasion literature to illustrate how information provision experiments can be implemented. Blesse (2021) studies whether tax complexity affects individuals' tax morale in Germany. Also using a representative sample in Germany, Doerrenberg & Peichl (2020) assess whether tax morale or reciprocity in information treatments affect tax compliance. Moreover, there is a growing literature on the role of emotions and the perceptions of tax authorities on intended tax compliance (Privitera et al., 2021; Olsen et al., 2018; Enachescu et al., 2019; Kasper et al., 2015). Finally, Górecki & Letki (2021) implement a large-scale cross-national survey experiment in 14 countries in Central-Eastern Europe to examine the effect of social norms.

4 Concluding Remarks and Open Questions

This paper focuses on survey methods and experimental extensions of conventional surveys. However, we want to emphasize that there is no superior method to measure and explain undeclared work or the shadow economy as a whole. Every method has strengths and weaknesses. Ideally, more than one method should be employed.

Survey techniques particularly lend themselves to study not only the size, but also the structure and determinants of the supply and demand of undeclared work. We elaborate on this in two ways. First, we examine how to measure and explain undeclared work with survey data. In Section 2, we summarize the empirical evidence from surveys conducted in Germany between 2001 and 2019. We find that self-reported participation in undeclared work has decreased since the early 2000s, whereas the potential remains constant at a higher level. The risk of being detected is perceived as generally high, but is lower in the latest Eurobarometer waves from 2013 and 2019. The fraction of respondents expecting severe sanctions for undeclared work has increased over time and with the magnitude of undeclared work. Regarding the perception of social norms, the fraction of respondents that assesses private demand for private undeclared work as absolutely unacceptable is the smallest compared to other types of undeclared work in all survey waves.

Second, following the overview of survey evidence for the case of Germany, we discuss five limitations of conventional survey data: excluding workers without permanent residence from the sample, focusing on the supply side, comparability over survey waves, social desirability bias in direct questions, and the lack of causal analysis. To remedy

these shortcomings, we introduce four experimental survey techniques in Section 3. We argue that the randomized response technique and list experiments are particularly useful for measuring the prevalence of undeclared work in surveys. Using conventional and experimental survey techniques would help cross-validate estimates. Information provision experiments and discrete choice experiments serve well to identify the causal effects of fiscal parameters or social norms on undeclared work. In contrast to the randomized response technique and list experiments that derive estimates on the aggregate level, information provision and discrete choice experiments gather participation rates on the individual level with exogenous variation of determinants. As such, they may contribute to a thorough understanding of the factors that lead individuals to participate in undeclared work and help design effective policies to combat it.

Applications of experimental survey techniques to questions of undeclared work are limited to few estimations of the prevalence in selected sectors and samples as well as investigations of the effectiveness of tax subsidies to encourage the demand of declared household services (see Section 3.2). This literature points to a number of avenues for further research using experimental survey techniques.

Using variants of the randomized response technique and list experiments would allow us to conduct robust validity checks of the estimates of undeclared work. Moreover, in information provision experiments and discrete choice experiments, a variety of determinants of undeclared work can be causally assessed. First, public policy variables should be examined more closely. These include fiscal parameters (e.g., tax rates or tax thresholds), deterrence measures (e.g., fines or detection probabilities), labor market regulations (e.g., minimum wages) or positive tax incentives (e.g., vouchers). Moreover, the effect of bureaucratic costs (e.g., costs that result from registering a worker) and the tax authorities' treatment of taxpayers that vary across countries and sectors should be examined.

It must be noted that (psychologically) complex concepts may be better studied with information provision experiments as they allow participants to concentrate on one experimental variation. This also holds for social norms that have been found to affect undeclared work and collaborative tax evasion (Feld & Larsen, 2012b; Burgstaller & Pfeil, 2022). Future experimental surveys could identify to what extent social norms determine undeclared work, distinguishing the type (empirical/ normative) and direction (honesty/ dishonesty) of a social norm.¹⁷ Since the activation of a social norm requires participants' receptiveness, the effect of social norms should be studied in information provision experiments that prime with a norm and observe variations in the willingness to participate in undeclared work or tax morale as a response to such priming. The suggested extensions of the current literature may foster our understanding of the oc-

¹⁷For a nuanced concept of social norms, see Bicchieri & Xiao (2009), and for the effect of empirical evasion expectations on collaborative tax evasion in the laboratory, see Burgstaller & Pfeil (2022).

currence of undeclared work and the effects of policy parameters on participation. They may further assist policy makers in adapting existing policy instruments or introducing novel measures to reduce undeclared work.

References

- Abraham, M., Lorek, K., Richter, F., & Wrede, M. (2017). Collusive Tax Evasion and Social Norms. *International Tax and Public Finance*, *24*(2), 179–197.
- Adriaenssens, S., Theys, T., Verhaest, D., & Deschacht, N. (2021). Subsidized Household Services and Informal Employment: The Belgian Service Voucher Policy. *Journal of Social Policy*, 1–22.
- Alesina, A., Stantcheva, S., & Teso, E. (2018). Intergenerational Mobility and Preferences for Redistribution. *American Economic Review*, *108*(2), 521–54.
- Alm, J., & Malézieux, A. (2021). 40 Years of Tax Evasion Games: A Meta-Analysis. *Experimental Economics*, *24*(3), 699–750.
- Antinyan, A., & Asatryan, Z. (2020). Nudging for Tax Compliance: A Meta-Analysis. *CESifo Working Paper, No. 8500*.
- Balafoutas, L., Beck, A., Kerschbamer, R., & Sutter, M. (2015). The Hidden Costs of Tax Evasion: Collaborative Tax Evasion in Markets for Expert Services. *Journal of Public Economics*, *129*, 14–25.
- Bansak, K., Hainmueller, J., & Hangartner, D. (2016). How Economic, Humanitarian, and Religious Concerns Shape European Attitudes Toward Asylum Seekers. *Science*, *354*(6309), 217–222.
- Bicchieri, C., & Xiao, E. (2009). Do the Right Thing: But Only if Others Do So. *Journal of Behavioral Decision Making*, *22*(2), 191–208.
- Bjørneby, M., Alstadsæter, A., & Telle, K. (2021). Limits to Third-Party Reporting: Evidence from a Randomized Field Experiment in Norway. *Journal of Public Economics*, *203*, 104512.
- Blair, G., & Imai, K. (2012). Statistical Analysis of List Experiments. *Political Analysis*, *20*(1), 47–77.
- Blair, G., Imai, K., & Zhou, Y.-Y. (2015). Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association*, *110*(511), 1304–1319.
- Blesse, S. (2021). Are Your Tax Problems an Opportunity Not to Pay Taxes? Evidence from a Randomized Survey Experiment. *ZEW Discussion Papers, No. 21-040*.

- Burgstaller, L., Doerr, A., & Necker, S. (2022). Do Household Tax Credits Increase the Demand for Legally Provided Services? *Unpublished Manuscript, Walter Eucken Institute.*
- Burgstaller, L., & Pfeil, K. (2022). You Don't Need an Invoice, Do You? An Online Experiment on Collaborative Tax Evasion. *Unpublished Manuscript, Walter Eucken Institute.*
- Clark, M. D., Determann, D., Petrou, S., Moro, D., & de Bekker-Grob, E. W. (2014). Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics, 32*(9), 883–902.
- Cohn, A., & Maréchal, M. A. (2016). Priming in Economics. *Current Opinion in Psychology, 12*, 17–21.
- Corbacho, A., Gingerich, D. W., Oliveros, V., & Ruiz-Vega, M. (2016). Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica. *American Journal of Political Science, 60*(4), 1077–1092.
- Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research, 40*(1), 169–193.
- Datta, N. (2019). Willing to Pay for Security: A Discrete Choice Experiment to Analyse Labour Supply Preferences. *CEP Discussion Paper, No. 1632.*
- De Bekker-Grob, E. W., Hol, L., Donkers, B., Van Dam, L., Habbema, J. D. F., Van Leerdam, M. E., ... Steyerberg, E. W. (2010). Labeled versus Unlabeled Discrete Choice Experiments in Health Economics: An Application to Colorectal Cancer Screening. *Value in Health, 13*(2), 315–323.
- De Gregorio, C., & Giordano, A. (2016). The Heterogeneity of Undeclared Work in Italy: Some Results from the Statistical Integration of Survey and Administrative Sources. *Rivista di Statistica Ufficiale*(2), 99–129.
- Doerr, A., Hartmann, C., & Sajons, C. (2022). Unregistered Work Among Refugees: Evidence from a List Experiment in Germany. *WWZ Working Paper, No. 2022/01.*
- Doerr, A., & Necker, S. (2021). Collaborative Tax Evasion in the Provision of Services to Consumers: A Field Experiment. *American Economic Journal: Economic Policy, 13*(4), 185–216.

- Doerrenberg, P., & Peichl, A. (2020). Tax Morale and the Role of Social Norms and Reciprocity. Evidence from a Randomized Survey Experiment. *FinanzArchiv: Public Finance Analysis*, 1–43.
- Droitcour, J., Caspar, R. A., Hubbard, M. L., Parsley, T. L., Visscher, W., & Ezzati, T. M. (2004). The Item Count Technique as a Method of Indirect Questioning: A Review of its Development and a Case Study Application. In *Measurement Errors in Surveys* (pp. 185–210). Wiley Online Library.
- Dybka, P., Kowalczyk, M., Olesiński, B., Torój, A., & Rozkrut, M. (2019). Currency Demand and MIMIC Models: Towards a Structured Hybrid Method of Measuring the Shadow Economy. *International Tax and Public Finance*, 26(1), 4–40.
- Ehler, I., Wolter, F., & Junkermann, J. (2021). Sensitive Questions in Surveys: A Comprehensive Meta-Analysis of Experimental Survey Studies on the Performance of the Item Count Technique. *Public Opinion Quarterly*, 85(1), 6–27.
- Elek, P., & Köllő, J. (2019). Eliciting Permanent and Transitory Undeclared Work from Matched Administrative and Survey Data. *Empirica*, 46(3), 547–576.
- Enachescu, J., Olsen, J., Kogler, C., Zeelenberg, M., Breugelmans, S. M., & Kirchler, E. (2019). The Role of Emotions in Tax Compliance Behavior: A Mixed-methods Approach. *Journal of Economic Psychology*, 74, 102194.
- European Commission. (2014). *Special Eurobarometer 402: Undeclared Work in the European Union*. European Commission.
- European Commission. (2020). *Special Eurobarometer 498: Undeclared Work in the European Union*. European Commission.
- Feld, L. P., & Larsen, C. (2012a). Self-Perceptions, Government Policies and Tax Compliance in Germany. *International Tax and Public Finance*, 19(1), 78–103.
- Feld, L. P., & Larsen, C. (2012b). *Undeclared Work, Deterrence and Social Norms: The Case of Germany*. Springer Science & Business Media.
- Feld, L. P., Necker, S., Pirsig, S., Kirbach, M., Pracz, K., & Petersen, H.-G. (2013). Evaluierung der Wirksamkeit der steuerlichen Förderung für Handwerksleistungen nach § 35a EStG. *Ernst & Young GmbH Wirtschaftsprüfungsgesellschaft*.

- Feld, L. P., & Schneider, F. (2010). Survey on the Shadow Economy and Undeclared Earnings in OECD Countries. *German Economic Review*, 11(2), 109–149.
- Feld, L. P., & Schneider, F. (2017). Reply to Gebhard Kirchgässner. *German Economic Review*, 18(1), 112–131.
- Feld, L. P., Schneider, F., & Baskaran, T. (2010). Tax Non-Compliance, Deterrence and Social Norms: Survey Evidence from Germany. *Unpublished Manuscript, Ruprecht-Karls-University Heidelberg*.
- Fergusson, L., Molina, C., & Riaño, J. F. (2019). Consumers as VAT "Evaders": Incidence, Social Bias, and Correlates in Colombia. *Economía*, 19(2), 21–68.
- Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2), 303–315.
- Fochmann, M., Fochmann, N., Kocher, M. G., & Müller, N. (2021). Dishonesty and Risk-Taking: Compliance Decisions of Individuals and Groups. *Journal of Economic Behavior & Organization*, 185, 250–286.
- Fox, J. A., & Tracy, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys*. SAGE Publications.
- Giles, D. E., Tedds, L. M., & Werkneh, G. (2002). The Canadian Underground and Measured Economies: Granger Causality Results. *Applied Economics*, 34(18), 2347–2352.
- Gingerich, D. W. (2010). Understanding Off-the-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys. *Political Analysis*, 18(3), 349–380.
- Glynn, A. N. (2013). What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. *Public Opinion Quarterly*, 77(S1), 159–172.
- Górecki, M. A., & Letki, N. (2021). Social Norms Moderate the Effect of Tax System on Tax Evasion: Evidence from a Large-Scale Survey Experiment. *Journal of Business Ethics*, 172(4), 727–746.
- Haaland, I., Roth, C., & Wohlfart, J. (2022). Designing Information Provision Experiments. *Journal of Economic Literature* (Forthcoming).

- Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating Vignette and Conjoint Survey Experiments Against Real-World Behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400.
- Hallsworth, M. (2014). The Use of Field Experiments to Increase Tax Compliance. *Oxford Review of Economic Policy*, *30*(4), 658–679.
- Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.
- Höglinger, M., & Jann, B. (2018). More Is Not Always Better: An Experimental Individual-Level Validation of the Randomized Response Technique and the Crosswise Model. *PloS ONE*, *13*(8), e0201770.
- Horiuchi, Y., Markovich, Z., & Yamamoto, T. (2021). Does Conjoint Analysis Mitigate Social Desirability Bias? *Political Analysis*, 1–15.
- Horodnic, I. A., & Williams, C. C. (2022). Tackling Undeclared Work in the European Union: Beyond the Rational Economic Actor Approach. *Policy Studies*, *43*(1), 21–55.
- Horodnic, I. A., Williams, C. C., Windebank, J., Zait, A., & Ciobanu, C. I. (2021). Explaining Consumer Motives to Purchase in the Informal Economy. *Plos One*, *16*(10), e0258686.
- Humburg, M., & Van der Velden, R. (2015). Skills and the Graduate Recruitment Process: Evidence from Two Discrete Choice Experiments. *Economics of Education Review*, *49*, 24–41.
- John, L. K., Loewenstein, G., Acquisti, A., & Vosgerau, J. (2018). When and Why Randomized Response Techniques (Fail To) Elicit the Truth. *Organizational Behavior and Human Decision Processes*, *148*, 101–123.
- Kasper, M., Kogler, C., & Kirchler, E. (2015). Tax Policy and the News: An Empirical Analysis of Taxpayers’ Perceptions of Tax-Related Media Coverage and its Impact on Tax Compliance. *Journal of Behavioral and Experimental Economics*, *54*, 58–63.
- Kirchgässner, G. (2017). On Estimating the Size of the Shadow Economy. *German Economic Review*, *18*(1), 99–111.

- Kirchner, A., Krumpal, I., Trappmann, M., & von Hermann, H. (2013). Measuring and Explaining Undeclared Work in Germany—An Empirical Survey with a Special Focus on Social Desirability Bias. *Zeitschrift für Soziologie*, *42*(4), 291–314.
- Korndörfer, M., Krumpal, I., & Schmukle, S. C. (2014). Measuring and Explaining Tax Evasion: Improving Self-Reports Using the Crosswise Model. *Journal of Economic Psychology*, *45*, 18–32.
- Krumpal, I. (2012). Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning. *Social Science Research*, *41*(6), 1387–1403.
- Krumpal, I. (2013). Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity*, *47*(4), 2025–2047.
- Krumpal, I., Jann, B., Korndörfer, M., & Schmukle, S. (2018). Item Sum Double-List Technique: An Enhanced Design for Asking Quantitative Sensitive Questions. *Survey Research Methods*, *12*(2), 91–102.
- Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997). Racial Attitudes and the “New South”. *The Journal of Politics*, *59*(2), 323–349.
- Kundt, T., Misch, F., & Nerré, B. (2013). Re-Assessing the Merits of Measuring Tax Evasions through Surveys: Evidence from Serbian Firms. *ZEW Discussion Papers*, No. 13-047.
- Lee, R. M. (1993). *Doing Research on Sensitive Topics*. Sage Publications.
- Lensvelt-Mulders, G. J., Hox, J. J., Van der Heijden, P. G., & Maas, C. J. (2005). Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods & Research*, *33*(3), 319–348.
- Li, J., & Van den Noortgate, W. (2019). A Meta-Analysis of the Relative Effectiveness of the Item Count Technique Compared to Direct Questioning. *Sociological Methods & Research*, 1–40.
- Lohse, T., & Simon, S. A. (2021). Compliance in Teams - Implications of Joint Decisions and Shared Consequences. *Journal of Behavioral and Experimental Economics*, *94*, 101745.
- Mas, A., & Pallais, A. (2017). Valuing Alternative Work Arrangements. *American Economic Review*, *107*(12), 3722–59.

- McFadden, D. (1986). The Choice Theory Approach to Market Research. *Marketing Science*, 5(4), 275–297.
- Meisters, J., Hoffmann, A., & Musch, J. (2020). Can Detailed Instructions and Comprehension Checks Increase the Validity of Crosswise Model Estimates? *PloS ONE*, 15(6), e0235403.
- Miller, J. D. (1984). A New Survey Technique for Studying Deviant Behavior. *The George Washington University. ProQuest Dissertations Publishing*. 8410488.
- OECD. (2021). Bringing Household Services Out of the Shadows: Formalising Non-Care Work In and Around the House. *OECD Publishing, Paris*.
- Olsen, J., Kasper, M., Enachescu, J., Benk, S., Budak, T., & Kirchler, E. (2018). Emotions and Tax Compliance Among Small Business Owners: An Experimental Survey. *International Review of Law and Economics*, 56, 42–52.
- Poulos, C., Yang, J.-C., Levin, C., Van Minh, H., Giang, K. B., & Nguyen, D. (2011). Mothers’ Preferences and Willingness to Pay for HPV Vaccines in Vinh Long Province, Vietnam. *Social Science & Medicine*, 73(2), 226–234.
- Privitera, A., Enachescu, J., Kirchler, E., & Hartmann, A. J. (2021). Emotions in Tax Related Situations Shape Compliance Intentions: A Comparison Between Austria and Italy. *Journal of Behavioral and Experimental Economics*, 92, 101698.
- Putnins, T., & Sauka, A. (2021). The Shadow Economy Index for the Baltic Countries, 2009–2020. *Research Paper, Stockholm School of Economics in Riga, Latvia*.
- Schneider, F. (2021). Do Different Estimation Methods Lead to Implausible Differences in the Size of the Non-Observed or Shadow Economies? A Preliminary Answer. *CESifo Working Paper, No. 9434*.
- Schneider, F., & Buehn, A. (2018). Shadow Economy: Estimation Methods, Problems, Results and Open Questions. *Open Economics*, 1, 1–29.
- Schneider, F., & Enste, D. H. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature*, 38(1), 77–114.
- Slemrod, J., & Weber, C. (2012). Evidence of the Invisible: Toward a Credibility Revolution in the Empirical Analysis of Tax Evasion and the Informal Economy. *International Tax and Public Finance*, 19(1), 25–53.

- Striegel, H., Ulrich, R., & Simon, P. (2010). Randomized Response Estimates for Doping and Illicit Drug Use in Elite Athletes. *Drug and Alcohol Dependence*, *106*(2-3), 230–232.
- Theys, T., Adriaenssens, S., Verhaest, D., Deschacht, N., & Rousseau, S. (2020). Disentangling Language from Ethnic Preferences in the Recruitment of Domestic Workers: A Discrete Choice Experiment. *Journal of Business Research*, *117*(1), 144–151.
- Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin*, *133*(5), 859–883.
- Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item Sum: A New Technique for Asking Quantitative Sensitive Questions. *Journal of Survey Statistics and Methodology*, *2*(1), 58–77.
- Trautmüller, R., Kijewski, S., & Freitag, M. (2019). The Silent Victims of Sexual Violence during War: Evidence from a List Experiment in Sri Lanka. *Journal of Conflict Resolution*, *63*(9), 2015–2042.
- Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. *American Economic Journal: Microeconomics*, *4*(4), 145–71.
- Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, *60*(309), 63–69.
- Williams, C. C., Bejaković, P., Mikulić, D., Franic, J., Kedir, A., & Horodnic, I. A. (2017). An Evaluation of the Scale of Undeclared Work in the European Union and Its Structural Determinants: Estimates Using the Labour Input Method. *European Commission*.
- Williams, C. C., & Kosta, B. (2021). Why Do Consumers Buy from Informal Sector Suppliers in East-Central Europe? A Case Study of Home Repair and Renovation Services. *Economic Alternatives*, *1*, 134–152.
- Williams, C. C., & Martinez-Perez, A. (2014). Why Do Consumers Purchase Goods and Services in the Informal Economy? *Journal of Business Research*, *67*(5), 802–806.

- Williams, C. C., & Nadin, S. (2014). Evaluating the Participation of the Unemployed in Undeclared Work: Evidence from a 27-Nation European Survey. *European Societies*, 16(1), 68–89.
- Williams, C. C., & Öz-Yalaman, G. (2021). Re-theorising Participation in Undeclared Work in the European Union: Lessons from a 2019 Eurobarometer Survey. *European Societies*, 23(3), 403–427.
- Wolter, F., & Diekmann, A. (2021). False Positives and the “More-is-Better” Assumption in Sensitive Question Research: New Evidence on the Crosswise Model and the Item Count Technique. *Public Opinion Quarterly*, 85(3), 836–863.
- Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis. *Metrika*, 67(3), 251–263.
- Žukauskas, V., & Schneider, F. (2016). Micro Based Results of Shadow Labour Market in the Baltic States, Poland, Sweden, and Belarus. *Applied Economics: Systematic Research*, 10(2), 117–134.