

Messung und Erklärung von Schwarzarbeit in Deutschland – Eine empirische Befragungsstudie unter besonderer Berücksichtigung des Problems der sozialen Erwünschtheit*

Measuring and Explaining Undeclared Work in Germany – An Empirical Survey with a Special Focus on Social Desirability Bias

Antje Kirchner

Institut für Arbeitsmarkt und Berufsforschung (IAB), Regensburger Str. 104, 90478 Nürnberg, Germany
antje.kirchner@iab.de

Ivar Krumpal

Universität Leipzig, Institut für Soziologie, Beethovenstraße 15, 04107 Leipzig, Germany
krumpal@sozio.uni-leipzig.de

Mark Trappmann

Institut für Arbeitsmarkt- und Berufsforschung (IAB), Regensburger Str. 104, 90478 Nürnberg, Germany
mark.trappmann@iab.de

Hagen von Hermanni

Universität Leipzig, Erziehungswissenschaftliche Fakultät, Karl-Heine-Straße 22b, 04229 Leipzig, Germany
hagen.von_hermanni@uni-leipzig.de

Zusammenfassung: Der vorliegende Beitrag geht der Frage nach, wie das Ausmaß von Schwarzarbeit in Deutschland im Rahmen von Befragungen der allgemeinen Bevölkerung möglichst valide geschätzt werden kann. In einem experimentellen Design wird die konventionelle direkte Befragungstechnik mit zwei Spezialtechniken, der Randomized-Response-Technik (RRT) und der Item-Count-Technik (ICT), verglichen. Die RRT und die ICT wurden für die Messung besonders heikler Verhaltensweisen entwickelt und sollen durch eine Erhöhung der Anonymität in der Interviewsituation sozial erwünschtes Antwortverhalten reduzieren. Unsere Befunde zeigen, dass die häufig angenommene Wirkung der beiden Spezialtechniken auf die Bereitschaft der Befragten, sozial unerwünschtes Verhalten zu berichten, nicht eindeutig ausfällt. Zudem werden theoretisch bedeutsame Einflussfaktoren von Schwarzarbeit diskutiert und deren Wirkung im Rahmen von multiplen Regressionsanalysen empirisch überprüft. Neben Gelegenheitsstrukturen sind vor allem soziale Normen gute Prädiktoren für die individuelle Entscheidung schwarzzuarbeiten.

Schlagworte: Schwarzarbeit; Schattenwirtschaft; Heikle Fragen; Soziale Erwünschtheit; Randomized-Response-Technik; Item-Count-Technik.

Summary: This article explores methods used to obtain a higher validity in estimates of the prevalence of undeclared work in Germany in surveys within the general population. Using an experimental design two “dejeopardizing” techniques are compared as alternatives to direct questioning when asking sensitive questions: the randomized response technique (RRT) and the item count technique (ICT). These techniques were specifically developed to reduce misreporting on sensitive topics: The goal is to elicit a higher proportion of honest answers from respondents by increasing the anonymity of the question-and-answer process. Our results suggest that neither RRT nor ICT provide unambiguous results with respect to more successful elicitation of reports of socially undesirable behavior. In addition, the theoretically significant influence of background variables is investigated empirically by means of multiple regression. Factors which foster illicit work are, aside from opportunity structures, social norms, which contribute significantly to the explanation of individual decisions to engage in undeclared work.

Keywords: Undeclared Work; Shadow Economy; Sensitive Questions; Social Desirability; Randomized Response Technique; Item Count Technique.

* Der Beitrag entstand im Rahmen einer Kooperation des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) und des Forschungsprojektes „Asking Sensitive Questions: Possibilities and Limits of Randomized Response and Other Techniques in Different Survey Modes“ (Teilprojekt Leipzig VO 684/11), das von der Deutschen Forschungsgemeinschaft im Rahmen des Schwerpunktprogramms 1292 „Survey Methodology“ gefördert wird.

1. Einleitung

Neuesten Schätzungen des Instituts für Angewandte Wirtschaftsforschung zufolge befindet sich die Schattenwirtschaft in Deutschland in den Jahren 2011 und 2012 auf einem Tiefstand, der zuletzt in den 1990er Jahren erreicht wurde (IAW 2013). Diese Studien zum Umfang der Schattenwirtschaft basieren meist auf indirekten makroökonomischen Schätzmethoden. Diese sind zwar in der Wissenschaft – ohne einen tragfähigen Konsens zu erreichen – vielfach debattiert, erweisen sich jedoch gegenüber den Modellen zugrunde liegenden Annahmen als sehr sensitiv (Thomas 1999; Thießen 2011). So reicht etwa die Spannweite der Schätzungen zum Ausmaß schattenwirtschaftlicher Aktivitäten in den Jahren 2000/01 von 1,3 bis 16,3 Prozent des Bruttoinlandproduktes (Feld et al. 2007). Die Tatsache, dass sowohl schattenwirtschaftliche Aktivitäten im Allgemeinen als auch Schwarzarbeit im Speziellen in den meisten Fällen unbeobachtet stattfinden (Koch 2005), kann hierbei kaum als zufriedenstellende Erklärung dienen. Auch eine trennscharfe Abgrenzung der verschiedenen Teilbereiche innerhalb der Schattenwirtschaft, die neben Schwarzarbeit auch den kriminellen Sektor umfasst, ist allein mit Ansätzen auf der Makroebene oft nur unzureichend möglich. Deshalb wird zur Messung der Größenordnung von Schwarzarbeit, als einem Teilbereich der Schattenwirtschaft, zusätzlich auf Ansätze auf der Mikroebene, wie beispielsweise Bevölkerungsbefragungen, zurückgegriffen (vgl. Merz & Wolff 1993; Mummert & Schneider 2001; Pedersen 2003; Feld & Larsen 2005, 2008; EC 2007).

Während Makroansätze zu Schätzungen führen, die oft als Obergrenze diskutiert werden, liefern Erhebungsmethoden auf der Mikroebene Schätzungen, die allgemein als Untergrenze gelten (Schneider 2003; Breusch 2005; Koch 2005; Feld et al. 2007; Pickard & Sarda 2011). Schließlich handelt es sich um ein sozial unerwünschtes und auch strafrechtlich relevantes Verhalten, das im Rahmen einer Bevölkerungsbefragung häufig geleugnet oder zumindest unterberichtet wird (Feld & Larsen 2008). Befragungsdaten besitzen gegenüber makroökonomischen Verfahren jedoch den entscheidenden Vorteil, dass neben reinen Prävalenzschätzungen auch multivariate Zusammenhänge auf individueller Ebene und somit auch die Motive von Individuen schwarzzuarbeiten, untersucht werden können. Voraussetzung für die Untersuchung dieser Zusammenhänge ist es jedoch, dass es keine systematischen Antworttendenzen gibt, die die Schätzungen verzerren (Hausman 2001).

Die vorliegende Studie verfolgt zwei Ziele: (1) die Messung und (2) die Erklärung von Schwarzarbeit auf der Grundlage von Individualdaten, die mithilfe von Verfahren erhoben wurden, die speziell für die Erhebung heikler Themen im Rahmen von Bevölkerungsbefragungen entwickelt worden sind (vgl. Tourangeau & Yan 2007; Krumpal 2013). Es handelt sich hierbei um die Randomized-Response-Technik (RRT; Warner 1965) und die Item-Count-Technik (ICT; Droitcour et al. 1991). Diese Verfahren erhöhen die Anonymität im Befragungsprozess und sollen so die Bedenken der Befragten zerstreuen, die eigene Schwarzarbeit im Rahmen einer Befragung zu berichten.

Damit trägt der Artikel wesentlich zum Schließen verschiedener Forschungslücken bei. *Zum einen* existieren bislang keine Bevölkerungsbefragungen zu Schwarzarbeit in der Bundesrepublik, in denen das Problem der sozialen Erwünschtheit adressiert wird. Im Rahmen der vorliegenden Studie soll daher das Potenzial spezieller Datenerhebungsverfahren zur Erhebung des heiklen Themas Schwarzarbeit empirisch untersucht und kritisch diskutiert werden (Boockmann et al. 2010: 100). Im Speziellen geht unser Beitrag der Frage nach, ob die RRT und die ICT sozial erwünschtes Antwortverhalten reduzieren und zu einer höheren Schätzung der Prävalenz von Schwarzarbeit führen als eine direkte Befragung („more is better“-Annahme; vgl. Tourangeau & Yan 2007). Hierbei stützen sich unsere empirischen Untersuchungen auf zwei bundesweite Bevölkerungsbefragungen aus dem Jahre 2010 ($n = 3.211$ bzw. $n = 1.603$), die auf einem experimentellen Design basieren und einen direkten Methodenvergleich erlauben. *Zweitens* sollen die eingesetzten Techniken auch im Hinblick auf ihre Effektivität in verschiedenen Populationen untersucht werden. Neben diesen uni- und bivariaten Analysen zur Prävalenz von Schwarzarbeit und einfachen Subgruppenvergleichen sollen *drittens* auch multiple Regressionsmodelle geschätzt werden. Diese erlauben es, aus theoretischen Überlegungen abgeleitete Hypothesen zur Erklärung von Schwarzarbeit empirisch zu überprüfen. So können etwa komplexe Zusammenhänge zwischen Schwarzarbeit und demographischen Merkmalen, Opportunitäts- und Anreizstrukturen sowie normativen Einstellungen auf der Individualebene genauer beleuchtet werden.

Der Artikel beginnt zunächst mit Überlegungen zur Definition und Abgrenzung der Konzepte ‚Schattenwirtschaft‘ und ‚Schwarzarbeit‘ (2.). Anschließend werden die RRT und die ICT als Techniken zur Messung heikler Merkmale vorgestellt (3.) und das Studiendesign der beiden experimentellen Sur-

veys beschrieben (4.). Nachdem auf deren Grundlage zunächst Prävalenzschätzungen berichtet werden (5.), werden Hypothesen zur Erklärung von Schwarzarbeit hergeleitet und mittels multipler Regressionsanalysen getestet (6.). Der Artikel endet mit einer Zusammenfassung der aus den Analysen gezogenen Schlussfolgerungen und einer Erörterung der weiteren Implikationen (7.).

2. Die Konzepte „Schattenwirtschaft“ und „Schwarzarbeit“

Analog zu den von der OECD (2002: 37ff) identifizierten Feldern unterscheiden Schneider & Enste (2007: 11) neben dem offiziellen Sektor *vier verschiedene Untersektoren* des inoffiziellen Sektors bzw. der Schattenwirtschaft im weiteren Sinn: Der *Haushaltssektor*, dem vor allem „Do-it-yourself“- und Eigenarbeit zugerechnet werden (Buehn et al. 2009), ist – wie der *informelle Sektor* auch – vor allem auf Selbstversorgungswirtschaft ausgerichtet. Zum informellen Sektor zählen Schneider & Enste insbesondere die Nachbarschaftshilfe und ehrenamtliche Tätigkeiten. Diese beiden Sektoren werden als legal bezeichnet, während die Schattenwirtschaft im engeren Sinn nur den irregulären und den kriminellen Sektor umfasst. Der *irreguläre Sektor* betrifft dabei im Kern die Schwarzarbeit. Diese bezeichnet insbesondere Meldeverstöße bezüglich Steuern und Sozialabgaben, die mangelnde Berechtigung zur Ausführung von Arbeiten sowie damit einhergehenden Leistungsmissbrauch (Schneider & Enste 2007: 11). Bei Evers (1987: 353) findet sich dieser Sektor unter dem Begriff informeller Sektor, gekennzeichnet durch eine in kleinen Einheiten angebotene marktorientierte Produktion bzw. Dienstleistungen. Dem *kriminellen Sektor* schließlich rechnen Schneider & Enste beispielsweise Schmuggel, Hehlerei oder auch Menschenhandel zu.¹

Unterschiede in den Definitionen und Abgrenzungen der Sektoren voneinander, vor allem auch im internationalen Vergleich, resultieren in unterschiedlichsten Schätzungen des Umfangs von Schwarzarbeit (vgl. Thomas 1999; Schneider et al. 2002;

¹ Neben der Unterscheidung von legalen und illegalen Tätigkeiten, wird häufig auch nach der Art der Transaktion unterschieden. Neben monetären basieren einige Aktivitäten auch auf nichtmonetären Transaktionen, wie beispielsweise Tauschhandel von (gestohlenen) Waren oder auch Nachbarschaftshilfe (Boockmann et al. 2010: 14). Einen Überblick über verschiedene Konzepte und Klassifikationen findet sich beispielsweise bei Schneider & Enste (2000).

Renooy et al. 2004; Koch 2005; Enste & Schneider 2006). Je nach Forschungsinteresse und -methode finden sich unterschiedliche Ansatzpunkte zur Definition der Begriffe Schattenwirtschaft und Schwarzarbeit, die sich an statistischen, steuerlichen oder rechtlichen Kriterien orientieren (Boockmann et al. 2010). Aufgrund der länderspezifischen Rechtsrahmen und Steuersysteme sind internationale Vergleiche daher immer mit Vorsicht anzustellen (Pedersen 2003).

In der BRD gestaltet sich das Definitionsproblem verhältnismäßig einfach, da seit 2004 ein „Gesetz zur Bekämpfung der Schwarzarbeit und illegalen Beschäftigung“ (SchwarzArbG) existiert, in dem eine explizite juristische Definition von Schwarzarbeit vorangestellt ist. Während in den meisten Ländern unter Schwarzarbeit ausschließlich Steuerhinterziehung und Sozialleistungsbetrug subsumiert werden (Boockmann et al. 2010), gehören gemäß der deutschen Gesetzeslage z. B. auch reiner Leistungsmissbrauch oder ausgewählte Verstöße gegen die Handwerksordnung zur Schwarzarbeit (§ 1 Abs. 2 SchwarzArbG).² Das SchwarzArbG enthält außerdem eine Negativdefinition, wonach solche Tätigkeiten nicht zur Schwarzarbeit zählen, die „nicht nachhaltig auf Gewinn gerichtet [sind und] gegen geringes Entgelt erbracht“ (§ 1 Abs. 3) werden. Damit sollen alle Aktivitäten straffrei bleiben, die „aus Gefälligkeit, im Wege der Nachbarschaftshilfe oder im Wege der Selbsthilfe“ (ebd.) erbracht werden, also laut OECD unter informelle oder Haushaltsproduktion fallen.

Eine einheitliche Definition, die sowohl alle rechtlichen Aspekte der Schwarzarbeit einschließt als auch eine saubere Abgrenzung von der informellen und Haushaltsproduktion zulässt und zudem Personen im Rahmen einer Befragung verständlich gemacht werden kann, erweist sich als Herausforderung. *Unter Schwarzarbeit werden im Folgenden jene produktiven Tätigkeiten verstanden, die erstens ein Arbeitseinkommen schaffen, zweitens an sich legal sind und drittens den Behörden nicht gemeldet werden, beispielsweise um Abgaben zu hinterziehen oder behördliche und gesetzliche Regeln wie Gesundheitsvorschriften oder Mindestlöhne zu umgehen* (Boockmann et al. 2010: 15; vgl. Renooy et al. 2004; EC 2007).

² Dieser und nachfolgende Verweise auf das „Gesetz zur Bekämpfung der Schwarzarbeit und illegalen Beschäftigung“ (SchwarzArbG) beziehen sich durchgängig auf dessen Fassung von 2004.

2.1 Ansätze zur Messung von Schattenwirtschaft und Schwarzarbeit

Die valide Messung des Umfangs und der Entwicklung von Schattenwirtschaft und Schwarzarbeit ist ebenso wie ihre Definition Gegenstand kontroverser Debatten (Thomas 1999; Pedersen 2003; Koch 2005; Enste & Schneider 2006; Thießen 2011). Dies liegt nicht zuletzt daran, dass die verschiedenen Methoden zu sehr unterschiedlichen Ergebnissen führen. Grundsätzlich können zwei Klassen von Messansätzen unterschieden werden: *indirekte* und *direkte Ansätze* (Pedersen 2003; Feld et al. 2007; Schneider & Enste 2007).

Zu den *indirekten Ansätzen*, die in der Regel Messungen auf der Makroebene beinhalten, zählen beispielsweise Diskrepanzanalysen, monetäre Schätzverfahren, physikalische Inputmethoden oder modellgestützte Verfahren (Boockmann et al. 2010: 63 ff.). Mithilfe dieser Ansätze ist es lediglich möglich, aggregierte Schätzungen zum Ausmaß der Schattenwirtschaft in einer Volkswirtschaft zu generieren (zumeist als Anteil am BIP). Für eines der zentralen Anliegen der wissenschaftlichen Forschung zur Schwarzarbeit, der Bestimmung von erklärenden Determinanten von Schwarzarbeit auf der Individualebene, sind diese Ansätze nicht nutzbar. Auf eine Erörterung dieser Verfahren wird daher an dieser Stelle verzichtet. Der interessierte Leser kann sich über diese Verfahren beispielsweise bei Feige (1990) oder bei Schneider & Enste (2000) informieren.

Direkte Verfahren auf der Mikroebene können dagegen Fragestellungen auf Branchen- oder Individualebene eher Rechnung tragen. Zudem ist hier eine klarere Differenzierung zwischen den Konstrukten Schattenwirtschaft und Schwarzarbeit möglich. Zu den direkten Mikroverfahren zählen neben Analysen administrativer Daten (der Finanzkontrolle Schwarzarbeit FKS) insbesondere Surveydaten. Während eine Nutzung administrativer Datenquellen meist nur eingeschränkt möglich und durch die hohe Dunkelziffer extrem selektiv ist (Feld et al. 2007: 8), können Bevölkerungsumfragen eine repräsentativere Schätzung zur Größenordnung von Schwarzarbeit liefern.³ Da es sich bei diesen Fragen um Prototypen heikler Fragen handelt – das zu berichtende Verhalten ist nicht nur sozial unerwünscht, sondern auch strafbar – ist davon

auszugehen, dass es dementsprechend stark unterberichtet oder komplett verschwiegen wird. Im äußersten Fall führt die Frage nach solchem Verhalten sogar zu Interviewabbrüchen (Tourangeau & Smith 1996: 276). Diesen Tendenzen versucht man daher mit speziellen Techniken entgegenzuwirken, die von Vertraulichkeitszusicherungen zu Beginn eines Interviews über verschiedene Formulierungsstrategien bis hin zu speziellen Fragetechniken reichen, bei denen eine probabilistische Beziehung zwischen der heiklen Frage und der individuellen Antwort hergestellt wird.

2.2 Bevölkerungsbefragungen als Instrumente zur Messung von Schwarzarbeit

Befragungen liefern bisher sehr unterschiedliche Schätzungen zum Umfang der Schwarzarbeit in Deutschland. Der Anteil derjenigen, die von sich berichteten schwarzgearbeitet zu haben, reicht von 3,0 Prozent der über 15-Jährigen im Eurobarometer 2007 (EC 2007: 96) bis zu 11,1 bzw. 7,2 Prozent der Befragten im Alter von 18 bis 74 Jahren in den Befragungen von Feld und Larsen in den Jahren 2005 und 2006 (Feld & Schneider 2010: 123).

Die unterschiedlichen Ergebnisse sind nicht zuletzt bedingt durch die Wahl verschiedener Designs (Stichprobe, Erhebungsmodus und -methoden) und Operationalisierungen. In den *Studien der Rockwool Foundation* für Deutschland (Pedersen 2003; Feld & Larsen 2005) findet sich beispielsweise eine sehr weite Definition von Schwarzarbeit, während andere Studien den Befragten vorab keine Definition geben (Schneider et al. 2002; Schneider 2008). Das *Eurobarometer 2007* verwendet für alle Länder die Definition der OECD von 2002, die ebenfalls relativ weit gefasst ist (Boockmann et al. 2010: 19). Schneider et al. (2002: 44) verwenden in der Formulierung der Items zur Messung von Schwarzarbeit den sogenannten Jedermann-Ansatz (vgl. Barton 1958). Hier wird Schwarzarbeit explizit genannt, allerdings als etwas Normales dargestellt, bagatellisiert und den Befragten somit suggeriert, dass es sich dabei um ein verbreitetes Phänomen handelt. Neuere experimentelle Studien zeigen jedoch, dass Variationen in der Formulierung einer heiklen Frage (z. B. entschuldigende Frageformulierung) keine nennenswerten Effekte auf individuelles Antwortverhalten haben und somit das Problem der sozialen Erwünschtheit nicht entschärfen können (Krumpal & Näher 2012).

Grundsätzlich können systematische Antworttendenzen („underreporting“) oder Item-Nonresponse

³ Mittels Bevölkerungsbefragungen wird in der Regel wiederum nur ein Teilbereich der Schwarzarbeit beleuchtet. Der Materialeinsatz oder -verbrauch für Schwarzarbeit beispielsweise geht daraus meist nicht hervor.

aufgrund sozialer Erwünschtheit zu einer geringeren Validität der Prävalenzschätzungen dieses Verhaltens führen („social desirability bias“; vgl. Tourangeau & Yan 2007). Insbesondere dann, wenn die Befragten nicht davon überzeugt sind, dass ihre Angaben vertraulich behandelt und anonymisiert ausgewertet werden, führen Bedenken hinsichtlich sozialer oder rechtlicher Sanktionen zur Untertreibung oder Verheimlichung des peinlichen oder illegalen Verhaltens (Fox & Tracy 1986). Je heikler und bedrohlicher Fragen von Befragten eingeschätzt werden, desto wahrscheinlicher ist es, dass systematische Antworttendenzen auftreten (Bradburn et al. 2004: 80; Lensvelt-Mulders et al. 2005a). Es stellt sich also die Frage, wie das Thema Schwarzarbeit in einer Bevölkerungsbefragung in geeigneter Weise umzusetzen ist, um ehrlichere Selbstauskünfte zu erhalten und somit eine validere Messung zu erzielen.

3. Ansätze zur Messung von sensitiven Merkmalen

Einen alternativen Ansatz zur Messung sensitiver Merkmale stellen in den letzten Jahrzehnten entwickelte Verfahren dar. Diese beruhen auf dem Prinzip, dem Befragten Anonymität bei der Beantwortung ihm möglicherweise unangenehmer Fragen nicht nur zuzusichern, sondern tatsächlich eine anonyme Befragungssituation zu schaffen, in der niemand aus einer gegebenen individuellen Antwort eindeutig auf das Verhalten des Befragten schließen kann (Lee 1993: 82). Zu diesen Verfahren zählen insbesondere die *Randomized-Response-Technik* (Warner 1965) und die *Item-Count-Technik* (Droitcour et al. 1991). Bei beiden Verfahren ist es auf der Grundlage wahrheitstheoretischer Überlegungen möglich, Prävalenzraten für Schwarzarbeit zu schätzen. Ein Überblick über beide Techniken sowie Alternativen findet sich bei Tourangeau & Yan (2007).

3.1 Die Randomized-Response-Technik

Über das Prinzip zufallsverschlüsselter Antworten wird mittels der Randomized-Response-Technik (RRT) Anonymität in der Befragungssituation hergestellt (Deffaa 1982). Warners (1965) ursprüngliche Methode wurde inzwischen in unzähligen Varianten weiterentwickelt und implementiert (Horvitz et al. 1967; Greenberg et al. 1969; Boruch 1971; Greenberg et al. 1971; Moors 1971; Kuk 1990; Mangat & Singh 1990; Mangat 1994). Allen

RRT-Formaten ist gemein, dass über einen Zufalls-generator eine probabilistische Beziehung zwischen der heiklen Frage und einer gegebenen Antwort hergestellt wird. Interviewer und Datennutzer können aus einer gegebenen Antwort nicht auf das Vorliegen bzw. Nichtvorliegen des sensitiven Merkmals bei einem Befragten schließen. In der Theorie wird angenommen, dass der Befragte die so hergestellte Anonymität wertschätzt und ehrlicher antwortet.

Im hier verwendeten „Forced Choice Design“ (Boruch 1971)⁴ wird ein Zufalls-generator (z. B. die letzten Ziffern der Seriennummer eines Geldscheins, ein Würfel oder Münzen) mit bekannter Wahrscheinlichkeitsverteilung verwendet, wobei der Befragte nach genau spezifizierten Regeln antworten soll. In der vorliegenden Studie wurden für die Randomisierung der Antworten drei Münzen verwendet. Hierbei wurden die Befragten gebeten, auf die heiklen Fragen gemäß den folgenden Regeln zu antworten:

Alle der folgenden Fragen lassen sich mit Ja oder Nein beantworten. Vor jeder Frage werfen Sie bitte die drei Münzen. Sagen Sie mir aber nicht, was Sie geworfen haben! Je nach Ergebnis des Münzwurfs, antworten Sie bitte wie folgt (...): Bei 3× Zahl, antworten Sie bitte immer mit „ja“; bei 3× Kopf, antworten Sie bitte immer mit „nein“; bei einem gemischten Ergebnis, also einer Kombination aus Zahl und Kopf, z. B. 2× Zahl und 1× Kopf, antworten Sie bitte immer wahrheitsgemäß.

Sie sehen, der Zufall entscheidet, ob Sie die Frage beantworten oder eine vorgegebene Antwort geben. Ihre Privatsphäre bleibt dadurch immer geschützt. Ich als Interviewer kenne das Ergebnis Ihres Münzwurfs nicht. Deshalb kann ich auch nie wissen, warum Sie „ja“ oder „nein“ sagen. Haben Sie das Münzverfahren verstanden? (...)

Nur der Befragte kennt das Ergebnis des Münzwurfs, so dass eine „ja“-Antwort keine eindeutigen Rückschlüsse auf das Verhalten des Befragten zulässt. Da aber die Wahrscheinlichkeitsverteilung des Münzwurfs bekannt ist, kann der Populationsanteil der Befragten, die die sensitive Frage mit „ja“ beantwortet haben, zusammen mit Standardfehlern und Konfidenzintervallen unverzerrt geschätzt werden: Der beobachtete Stichprobenanteil der „Ja“-Antworten $\hat{\Phi}$ dient hierbei als Schätzer für den unbekanntes Gesamtanteil der „Ja“-Antworten in der Population Φ . Diese setzen sich aus dem aufgrund des Münzwurfs erzwungenen Anteils „ja“-Antworten (hier: $p_1 = 0,125$) sowie aus dem Anteil ehrlicher „Ja“-Antworten ($p_2 = 0,75$) gewichtet mit

⁴ Lensvelt-Mulders et al. (2005b) vergleichen verschiedene RRT-Varianten und kommen zu dem Schluss, dass die „Forced Choice“-Variante unter den RRT-Designs die höchste statistische Effizienz aufweist.

dem (unbekannten) Anteil π des sensitiven Verhaltens wie folgt zusammen: $\Phi = p_1 + p_2 * \pi$. Nach einer entsprechenden Umstellung der Gleichung, ergibt sich dann der Anteilsschätzer $\hat{\pi}_{\text{RRT}}$ (Lensvelt-Mulders et al. 2005b):

$$\hat{\pi}_{\text{RRT}} = \frac{\hat{\Phi} - p_1}{p_2} \quad (1)$$

Die Stichprobenvarianz von $\hat{\pi}_{\text{RRT}}$ lässt sich wie folgt berechnen:

$$\text{Var}(\hat{\pi}_{\text{RRT}}) = \frac{\hat{\Phi} * (1 - \hat{\Phi})}{n * (p_2)^2} \quad (2)$$

Dem Vorteil der Anonymitätssteigerung stehen erhöhte Erhebungskosten (z. B. größere Stichproben, längere Interviewdauern und besondere Interviewerschulungen) sowie eine höhere kognitive Belastung des Befragten gegenüber (Lensvelt-Mulders & Boeije 2007). Allerdings kann Letzteres durch eine genaue und saubere Design- und Instrumentenentwicklung reduziert werden (vgl. Krumpal 2010). Da durch die Zufallsverschlüsselung die statistische Unsicherheit steigt, sind im Vergleich zu direkten Befragungen größere Stichproben notwendig, um das gleiche Niveau statistischer Power zu erreichen.

3.2 Die Item-Count-Technik

Als *Item-Count-Technik* (Droitcour et al. 1991)⁵ wird eine indirekte Befragungsmethode bezeichnet, die keine Schlüsse auf individuelle Antworten der Befragten zulässt, aber einen Schätzer für die Gesamtprävalenz in der untersuchten Population generiert. Das heikle binäre Item wird zusammen mit einer Reihe anderer, nicht heikler binärer Items abgefragt. Die Befragten werden gebeten, nur die Anzahl der zutreffenden Items zu berichten, anstatt alle Items einzeln zu beantworten. Zur Schätzung der Prävalenz des heiklen Ziel-Items, setzt die ICT eine Unterteilung in mindestens zwei homogene Teilstichproben voraus. Während Befragte der ersten Teilstichprobe eine Liste von nicht heiklen Items beantworten („short list“; SL), erhalten Befragte der zweiten Teilstichprobe die exakt gleiche Liste und zusätzlich das heikle Item („long list“; LL). Eine Schätzung der Prävalenz ($\hat{\pi}_{\text{ICT}}$) des heiklen Items lässt sich nun als Differenz der Mittelwerte der beiden Teilstichproben (Mittelwert der Long

List \bar{x}_{LL} minus Mittelwert der Short List \bar{x}_{SL}) bestimmen (Droitcour et al. 1991):

$$\hat{\pi}_{\text{ICT}} = \bar{x}_{\text{LL}} - \bar{x}_{\text{SL}} \quad (3)$$

Die Stichprobenvarianz des Schätzers ergibt sich aus der Summe der Varianzen der beiden Mittelwerte:

$$\text{Var}(\hat{\pi}_{\text{ICT}}) = \text{Var}(\bar{x}_{\text{LL}}) + \text{Var}(\bar{x}_{\text{SL}}) \quad (4)$$

Eine offensichtliche Schwachstelle der ICT ist die relativ große Varianz ihres Schätzers. Diese steigt sowohl mit der Varianz der einzelnen nicht heiklen Items als auch mit deren Anzahl und erfordert somit eine dementsprechend höhere Fallzahl als eine vergleichbare Direktbefragung (Tsuchiya 2007). Da in der oben dargestellten Single-List-Variante der ICT das heikle Item lediglich in einer Teilstichprobe erhoben wird, sinkt (halbiert sich) zudem die Stichprobengröße, was zu einer weiteren Vergrößerung der Varianz des Anteilsschätzers führt (Droitcour et al. 1991). Um die statistische Effizienz des ICT-Schätzers zu steigern, kann auf die sogenannte Double-List-Variante der ICT zurückgegriffen werden: Bei diesem Design wird das heikle Ziel-Item in beiden Gruppen erhoben. Dazu werden in den beiden Teilstichproben jeweils zwei verschiedene Listen in zwei aufeinanderfolgenden Frageblöcken eingesetzt. Dabei erhält die erste Teilstichprobe im ersten Frageblock die lange Liste mit dem heiklen Item (long list 1; LL1), während die zweite Teilstichprobe die korrespondierende kurze Liste ohne das heikle Item erhält (short list 1; SL1). Im zweiten Frageblock werden dagegen Befragte der ersten Teilstichprobe aufgefordert, auf eine kurze Liste ohne die heikle Frage zu antworten (diese enthält andere Items als die kurze Liste im ersten Frageblock; short list 2; SL2). Befragte in der zweiten Teilstichprobe erhalten im zweiten Frageblock die korrespondierende lange Liste (long list 2; LL2), die wiederum dasselbe heikle Item (SWA) enthält, das im ersten Frageblock von der ersten Teilstichprobe beantwortet wurde. Tabelle 1 veranschaulicht das gewählte ICT-Double-List-Design.

Droitcour et al. (1991) empfehlen die Verwendung von drei bis fünf nicht heiklen Items pro Liste, um eine hinreichend hohe Anonymität herzustellen und gleichzeitig die kognitive Belastung moderat zu halten. Diese Empfehlung hat sich in empirischen Studien bewährt (Ahart & Sackett 2004; Coutts & Jann 2011) und wurde deshalb auch in der vorliegenden Studie umgesetzt. Um keine zu großen Standardfehler zu erhalten, wurden in der Implementation unserer ICT-Studie jeweils drei nicht sensitive Items verwendet (je länger die „short list“, desto größer die Varianz).

⁵ Manchmal wird diese Technik auch als „block total response“ (Smith et al. 1974; Raghavarao & Federer 1979) als „unmatched count technique“ (Dalton et al. 1994; Ahart & Sackett 2004) oder „list experiment“ (Kuklinski et al. 1997) bezeichnet.

Tabelle 1 ICT-Double-List-Design

	Teilstichprobe 1	Teilstichprobe 2
Frageblock 1	LL1: SL1 & SWA	SL1
Frageblock 2	SL2	LL2: SL2 & SWA

Eine Übersicht zur Analyse dieser Double-List-ICT findet sich in Droitcour et al. (1991) und Biemer et al. (2005). Beim *Double-List-Design* erhält man basierend auf den beiden Frageblöcken zwei separate Schätzungen des heiklen Verhaltens

$$\hat{\pi}_1 = \bar{x}_{LL1} - \bar{x}_{SL1} \quad (5)$$

und

$$\hat{\pi}_2 = \bar{x}_{LL2} - \bar{x}_{SL2} \quad (6)$$

Deren Mittelwert gibt wiederum die Gesamtprävalenz an (Coutts et al. 2011):

$$\begin{aligned} \hat{\pi}_{DL} &= \frac{\hat{\pi}_1 + \hat{\pi}_2}{2} \\ &= \frac{(\bar{x}_{LL1} - \bar{x}_{SL1}) + (\bar{x}_{LL2} - \bar{x}_{SL2})}{2} \\ &= \frac{(\bar{x}_{LL1} - \bar{x}_{SL2}) + (\bar{x}_{LL2} - \bar{x}_{SL1})}{2} \end{aligned} \quad (7)$$

Weiterhin ergibt sich im Double-List-Design ein konservativer und im Vergleich zum Single-List-Design effizienterer Schätzer der Stichprobenvarianz (Coutts et al. 2011):

$$\begin{aligned} \text{Var}(\hat{\pi}_{DL}) &= \frac{\text{Var}(\hat{\pi}_1) + \text{Var}(\hat{\pi}_2) + 2\text{Cov}(\hat{\pi}_1, \hat{\pi}_2)}{4} \\ &= \frac{\text{Var}(\bar{x}_{LL1} - \bar{x}_{SL2}) + \text{Var}(\bar{x}_{LL2} - \bar{x}_{SL1})}{4} \end{aligned} \quad (8)$$

3.3 Bisherige empirische Evidenz zur ICT und RRT

Nur sehr selten sind Forscher in der Lage, die Anteilswerte für sensitive Items, die mithilfe der unterschiedlichen Verfahren geschätzt werden, direkt zu validieren. Da sozial unerwünschtes Verhalten bei direkter Befragung unterberichtet wird, unterstellt man in Abwesenheit von Validierungsdaten, dass dasjenige Verfahren das sensitive Verhalten besser misst, das zu einem höheren Anteilswert führt („more is better“-Annahme).

Verschiedene experimentelle Studien, die die Ergebnisse der RRT oder der ICT mit alternativen Methoden der Datenerhebung vergleichen, kommen vielfach zu dem Schluss, dass die Verwendung der RRT bzw. der ICT im Vergleich zu direkten Befragungen zu höheren und damit valideren Schätzungen von illegalen, stigmatisierenden oder sozial unerwünschten Verhaltensweisen oder Einstellungen

führt (Weissman et al. 1986; Dalton et al. 1994; Kuklinski et al. 1997; Wimbush & Dalton 1997; LaBrie & Earleywine 2000; van der Heijden et al. 2000; Lara et al. 2004; Lensvelt-Mulders et al. 2005a; Lara et al. 2006; Tsuchiya et al. 2007; Krumpal 2010; Holbrook & Krosnick 2010a; Krumpal 2012). Andere Studien wiederum konnten die der RRT oder der ICT zugeschriebenen Erfolge nicht replizieren und berichten gemischte Ergebnisse (Droitcour et al. 1991; Umesh & Peterson 1991; Ahart & Sackett 2004; Biemer et al. 2005; Tourangeau & Yan 2007; Tsuchiya & Hirai 2010; Holbrook & Krosnick 2010b; Coutts & Jann 2011).

Zu den Vorzügen beider Verfahren zählt die höhere Anonymität im Befragungsprozess. Allerdings sind sowohl die RRT als auch die ICT, verglichen mit der direkten Befragung, statistisch weniger effizient, wobei die ICT die geringste statistische Effizienz aufweist (Corstange 2009; Holbrook & Krosnick 2010a, 2010b; Coutts & Jann 2011). Zudem führen beide Methoden zu einem erhöhten kognitiven und zeitlichen Aufwand beim Befragten, was wiederum höhere Entwicklungs- und Datenerhebungskosten nach sich zieht.

Insgesamt betrachtet sind die empirischen Ergebnisse nicht immer eindeutig, welche der beiden Methoden zu bevorzugen ist (vgl. Holbrook & Krosnick 2010b). Es zeigt sich jedoch, dass der Einsatz der RRT oder der ICT sozial unerwünschtes Antwortverhalten häufig reduzieren und damit die Validität der Messung sensitiver Merkmale erhöhen kann. In Anknüpfung an die bisherige Forschung gehen wir in der vorliegenden Arbeit der Frage nach, ob die beiden Methoden jeweils zu valideren Ergebnissen führen als eine konventionelle direkte Befragung und ob sich dies unter der „more is better“-Annahme auch in der Schätzung höherer Prävalenzen von ausgeübter Schwarzarbeit manifestiert.

4. Die Studie: Stichprobe und Vorgehensweise

Die empirischen Untersuchungen stützen sich auf zwei deutschlandweite experimentelle CATI-Studien: Von zwei verschiedenen Erhebungsinstituten (ForschungsWerk Nürnberg; USUMA Berlin), die

mit je einer der Studien betraut waren, wurden im Zeitraum von Oktober bis Dezember 2010 in der RRT-Studie 3.211 Interviews (approximativ AAPOR Response Rate 1; 17,5 %, AAPOR 2011) und in der ICT-Studie 1.603 Interviews geführt (RR1 15,2 %, AAPOR 2011).

Die RRT-Studie setzt sich aus zwei Zufallsstichproben aus den Registern der Bundesagentur für Arbeit (BA) zusammen. Bei der ersten Stichprobe handelt es sich um Personen im Alter von 18 bis 70 Jahren, die am 31. Dezember 2009 einer abhängigen Beschäftigung nachgegangen sind. Es handelt sich dabei um Personen, die Tätigkeiten nachgehen, bei denen Sozialversicherungsbeiträge abgeführt werden, unabhängig vom Umfang oder der Art dieser Tätigkeit. Selbständige Personen und Beamte sind von der Analyse ausgeschlossen, da diese keine Sozialversicherungsbeiträge bezahlen müssen. Personen, die einer gemeldeten geringfügigen Beschäftigung nachgehen, sind dagegen in der Auswahlgrundlage enthalten. Die zweite Stichprobe umfasst Personen im Alter von 18 bis 64 Jahren, die am 30. Juni 2010 Arbeitslosengeld II (ALG II) erhalten haben.⁶

Die Stichprobe der ICT-Studie wurde über die ADM-Telefonstichprobe „Easy Sample“ ausgewählt (Häder & Gabler 1998). Gezogen wurden Festnetz-Telefonnummern in ganz Deutschland proportional zur Einwohnerstruktur und regional geschichtet nach Kreisen und Ortsgrößenklassen. Über ein RDD („random digit dialing“) Verfahren wurde sichergestellt, dass auch nicht eingetragene Festnetzanschlüsse Teil der Auswahlgesamtheit waren. Anhand eines Schwedenschlüssels wurden in einem zweiten Schritt die Zielpersonen unter den Haushaltmitgliedern zwischen 18 und 70 Jahren zufällig ermittelt.

Bei der Zuweisung zu den Experimentalgruppen wurde die relativ geringere statistische Effizienz der RRT- und ICT-Schätzer berücksichtigt, indem jeweils etwa doppelt so viele Fälle in diesen Modus randomisiert wurden als in die direkte Befragung (Cohen 1988): In der RRT-Studie fanden von den insgesamt 3.211 realisierten Interviews 1.145 in der direkten Experimentalkondition statt, während 2.066 Befragte mittels RRT befragt werden sollten. Allerdings kam es in der RRT-Studie in nicht vernachlässigbarem Umfang zu Verweigerungen der RRT-Methode. Insgesamt 274 Personen (13,3 %) hatten keine Münze parat oder wollten aus anderen

Gründen die Frage nicht mit der RRT beantworten. Diese Gruppe unterläuft damit die Randomisierung. Im Rahmen der Studie wurden diese Verweigerer (im Folgenden auch „Wechsler“ genannt) direkt befragt und in allen Analysen getrennt ausgewiesen. In der ICT-Studie wurden von 1.603 Interviews 500 in der direkten Kondition und 1.103 in der Experimentalkondition mit dem Double-List-Verfahren erhoben.

Beide Studien waren in ihrem Fragebogendesign und den Formulierungen identisch, sie unterschieden sich lediglich in den Experimentalkonditionen. Das Rahmenthema der Studie „Lebens- und Arbeitswelten in Deutschland“ erlaubte es, zunächst Fragen zum individuellen Erwerbsverlauf oder auch zum Freizeitverhalten der Befragten zu stellen. Erst nach diesen Themenblöcken folgte der eigentliche Experimentaltel beider Studien, sodass die heiklen Fragen zu Schwarzarbeit thematisch in einen stimmigen Kontext eingebettet wurden. Die RRT wurde dabei wie in Abschnitt 3.1 dargestellt implementiert. Für die ICT wurde die in Abschnitt 3.2 beschriebene Double-List-Variante umgesetzt (für den Wortlaut der Instruktionen beider Experimentaltel siehe den Online-Anhang unter www.zfs-online.org). An die bisherige Forschungspraxis anknüpfend (Pedersen 2003; Feld & Larsen 2005; EC 2007; Schneider & Enste 2007; Williams 2009; Boockmann et al. 2010) sollte durch das Voranstellen einer Definition von Schwarzarbeit ein einheitliches Verständnis sichergestellt werden. Die verwendete Definition von Schwarzarbeit lautet wie folgt:

Unter Schwarzarbeit verstehen wir Tätigkeiten, die ein Arbeitseinkommen schaffen, den staatlichen Behörden aber nicht gemeldet werden, etwa um die damit verbundene Pflicht zur Entrichtung von Steuern und Sozialabgaben zu umgehen.

Kriminelle Aktivitäten, wie bspw. Drogenanbau oder -handel, zählen hingegen NICHT zur Schwarzarbeit.

Anschließend wurden die Items zur Schwarzarbeit erhoben. Die Interviewer wurden speziell in der jeweiligen Technik geschult und die Qualität der Interviews wurde stichprobenartig überwacht.

In beiden Studien wurde in Anlehnung an andere Studien (vgl. dazu EC 2007; Pfau-Effinger 2009) nach dem Abnehmer der Schwarzarbeit differenziert, also Schwarzarbeit für eine Privatperson und für eine Firma getrennt erhoben. Tabelle 2 gibt den Wortlaut der erhobenen Items wieder und macht die Grenzen einer Bevölkerungsumfrage deutlich. Schwarzarbeit von Firmen für Firmen kann beispielsweise auf diese Art nicht sinnvoll erhoben werden.

⁶ Für beide Stichproben wurden jeweils diejenigen Stich-tage mit dem aktuellsten Datum als Auswahlgrundlage gewählt.

Tabelle 2 Items zur Messung von Schwarzarbeit

Item		Studie	Gültige Antworten	Item- Nonresponse	N
Privatperson	Haben Sie in diesem Jahr für eine Privatperson schwarz gearbeitet?	RRT	Direkt: 1.142	3	1.145
			RRT: 1.790	2	1.792
			Wechsler: 272	2	274
		ICT	Direkt: 498	2	500
		ICT: 1.103	0	1.103	
Firma	Haben Sie in diesem Jahr für eine Firma gearbeitet, die Ihnen Ihr Einkommen ganz oder teilweise bar, d. h. an den Behörden vorbei, ausgezahlt hat?	RRT	Direkt: 1.142	3	1.145
			RRT: 1.789	3	1.792
			Wechsler: 273	1	274
		ICT	Direkt: 497	3	500
		ICT: 1.103	0	1.103	

Diese Fragen, die in der RRT-Studie nur mit ja oder nein bzw. der Anzahl der Items in der ICT-Studie beantwortet werden konnten, sind die zentralen abhängigen Variablen in unseren empirischen Analysen. Abschließend wurden in beiden Studien inhaltliche und kontrollierende Hintergrundinformationen zu theoretisch bedeutenden Einflussfaktoren von Schwarzarbeit erhoben.⁷

5. Messung von Schwarzarbeit: Schätzungen der Prävalenz

Abbildung 1 zeigt auf der y-Achse die Punktschätzungen für den geschätzten Anteil derjenigen Personen, die für eine Privatperson schwarzarbeiten, sowie deren 95 %-Konfidenzintervalle auf der Grundlage der verschiedenen Experimentalkonditionen der ICT- bzw. der RRT-Studie (x-Achse). Entgegen den theoretischen Erwartungen liegen die mit der RRT (1,22 % und 4,07 %) oder der ICT (0,26 %) geschätzten Prävalenzen nicht immer höher als die der direkten Abfrage der RRT-Studie (1,77 % und 3,29 %) und der ICT-Studie (3,41 %). Die Ergebnisse der ICT und der Beschäftigtenstichprobe der RRT-Studie zeigen, dass die Prävalenzschätzungen in der Experimentalkondition sogar unter denen der direkten Abfrage (0,26 % vs. 3,41 % und 1,22 % vs. 1,77 %) liegen. Die Unterschiede zwischen den Experimentalgruppen und Kontrollgruppen sind für beide Studien nicht sig-

nifikant. Insgesamt betrachtet fällt auf, dass in der RRT-Studie Befragte der ALG-II-Stichprobe tendenziell insgesamt mehr Schwarzarbeit für eine Privatperson berichten als Befragte der RRT-Beschäftigtenstichprobe. Die Schätzungen aus der ICT-Studie, die auf einer allgemeinen Bevölkerungstichprobe basieren, liegen zwischen den für die RRT-Stichproben geschätzten Werten.

In Abbildung 2 sind analog zu Abbildung 1 die Punktschätzungen der verschiedenen Experimentalkonditionen der RRT- und der ICT-Studie für das Item „Schwarzarbeit für eine Firma“ mit den zugehörigen 95 %-Konfidenzintervallen abgebildet. In der ALG-II-Stichprobe erzielt die RRT auf dem 5 %-Niveau einen signifikant höheren Schätzwert als die Schätzung aus der direkten Erhebung (5,98 % vs. 1,56 %). In der Beschäftigtenstichprobe der RRT-Studie liegen dagegen keine signifikanten Unterschiede zwischen den Befragungskonditionen vor. Die mittels RRT geschätzte Prävalenz (-0,86 %) ist in dieser Stichprobe nicht nur geringer als die direkt erhobene (0,53 %), die RRT-Schätzung führt sogar zur Schätzung einer negativen Prävalenz. Die Erhebung mittels ICT führt für dieses Item zu einem auf dem 5 %-Niveau signifikant höheren Anteilswert von Schwarzarbeit als die Erhebung durch eine direkte Befragung (6,41 % vs. 1,21 %). Ähnlich wie bei dem Item „Schwarzarbeit für eine Privatperson“ berichten ALG-II-Empfänger der RRT-Studie mehr Schwarzarbeit für eine Firma als die Befragten der RRT-Beschäftigtenstichprobe. Die in der ICT-Studie ermittelten Ergebnisse sind für dieses Item denen der ALG-II-Stichprobe ähnlicher, während sie für das Item „Privatperson“ zwischen diesen und der Beschäftigtenstichprobe lagen.

Zusammengefasst lässt sich festhalten, dass weder in der RRT- noch der ICT-Experimentalkondition

⁷ So wurden neben soziodemographischen Merkmalen (Geschlecht, Alter, Bildung) beispielsweise die Stellung im Beruf, die lokale Vernetzung, Mitgliedschaften in Vereinen, Einstellungen zu Schwarzarbeit sowie die geschätzte Prävalenz von Schwarzarbeit im Umfeld der Befragungsperson erhoben.

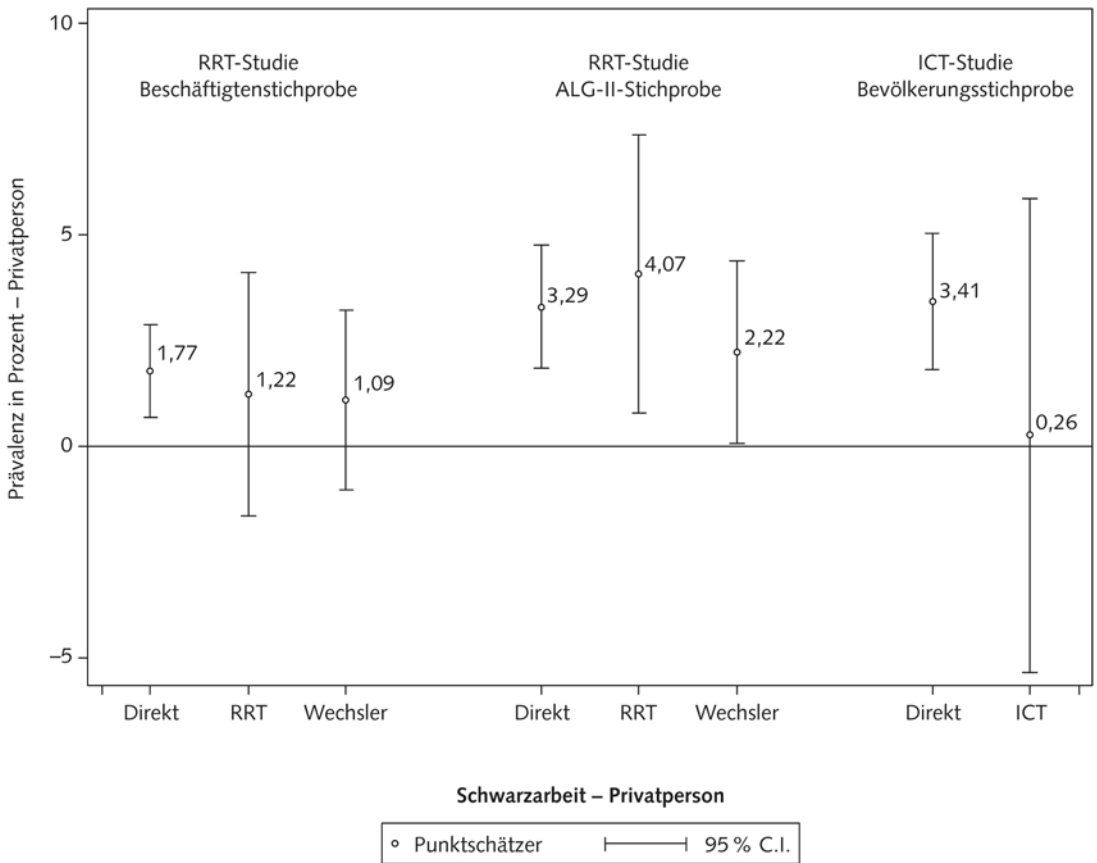


Abb. 1 Schwarzarbeit für eine Privatperson – Geschätzte Anteilswerte und 95 %- Konfidenzintervalle (in %)

konsistent die erwarteten höheren Prävalenzen beobachtet werden können. Angesichts dieser Ergebnisse stellt sich die Frage, warum in den Experimentalkonditionen in manchen Fällen die erwarteten höheren Prävalenzen beobachtet werden und in anderen nicht. Dass die Spezialverfahren bei der Schwarzarbeit für eine Firma besser funktionieren – hier finden sich die einzigen signifikanten Differenzen, beide in der erwarteten Richtung – als bei der Schwarzarbeit für eine Privatperson kann daran liegen, dass „Schwarzarbeit für eine Firma“ von den Befragten als das heiklere der beiden Items wahrgenommen wird, da dadurch eine quasi-institutionalisierte Form der Schwarzarbeit geschaffen wird. Grundsätzlich werden Verstöße gegen das SchwarzArbG als Ordnungswidrigkeit, aber auch als Straftat bewertet und gehen mit empfindlichen Geldbußen bis zu 300.000 Euro für Arbeitgeber, sowie bis zu fünf Jahren Haft für Arbeitnehmer und Arbeitgeber einher (§ 8 SchwarzArbG). Bei der Fest-

stellung des Strafmaßes wird dabei neben der „Schwere“ bzw. Quantität der verübten Schwarzarbeit auch unter Berücksichtigung der zu Grunde liegenden Absicht, d. h. ob es sich um das (unabsichtliche) bloße Nichtmelden oder um Vorsatz handelt, geurteilt (Zoll 2011; vgl. auch Feld et al. 2007: 12). In besonders schweren Fällen und insbesondere bei professionell organisierten Formen von Steuerhinterziehung können sogar Haftstrafen von bis zu 10 Jahren verhängt werden (§ 370a Abgabenordnung). Pfau-Effinger (2009: 89) argumentiert außerdem, dass Schwarzarbeit in Privathaushalten kulturell bedingt in Deutschland eher akzeptiert ist, seltener entdeckt und eher als Ordnungswidrigkeit geahndet wird als Schwarzarbeit in der „regulären“ Wirtschaft.

Für Personen im Arbeitslosengeld-II-Bezug wird zudem neben der Ahndung der Schwarzarbeit an sich (§ 9 SchwarzArbG), noch der Sozialleistungsbetrug gemäß § 263 StGB verfolgt, der mit einer Freiheits-

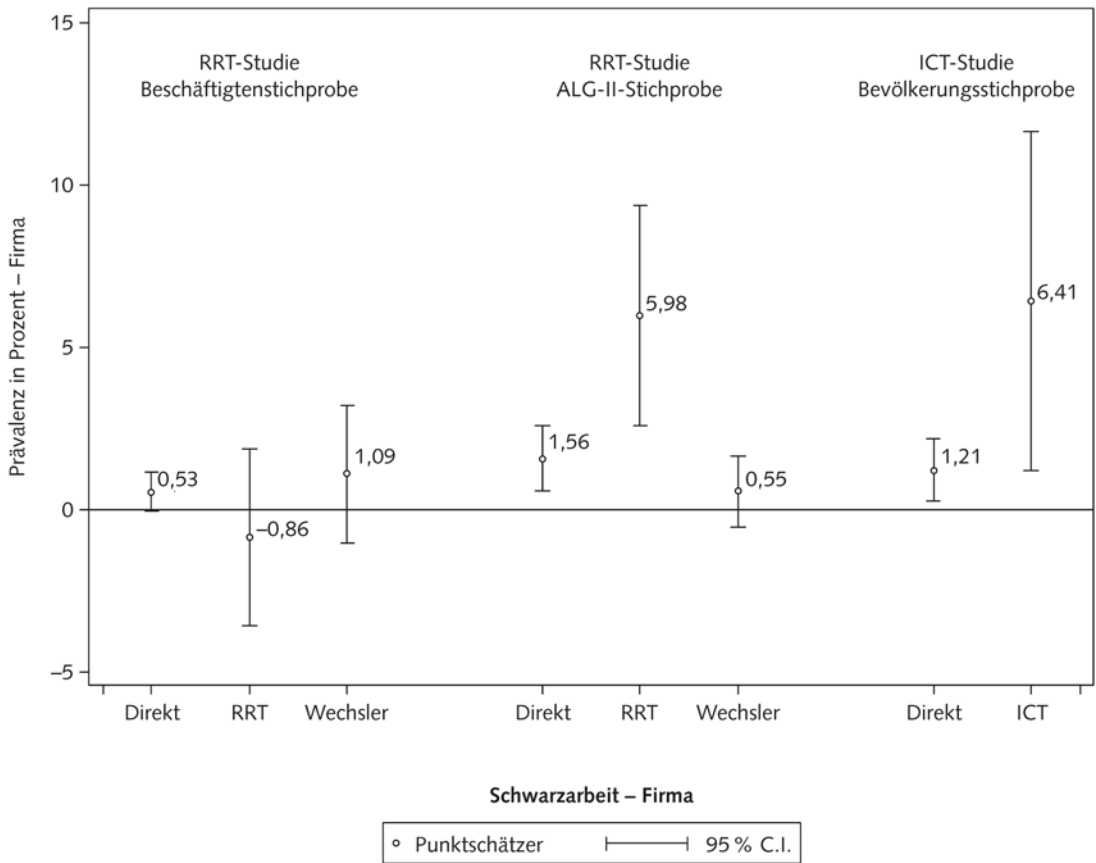


Abb. 2 Schwarzarbeit für eine Firma – Geschätzte Anteilswerte und 95 %-Konfidenzintervalle (in %)

strafe bis zu fünf Jahren oder mit einer Geldstrafe bestraft werden kann (Zoll 2011). Daher kann angenommen werden, dass die Items zur Schwarzarbeit in dieser Gruppe als noch heikler wahrgenommen werden. Argumentiert man nun basierend auf den Erkenntnissen anderer Studien (Lensvelt-Mulders et al. 2005a), dass diese indirekten Fragetechniken umso besser funktionieren, je heikler die betreffenden Items wahrgenommen werden, so würde man genau die beobachteten Resultate erwarten: Die Techniken für sensitive Fragen funktionieren bei ALG-II-Empfängern besser als bei Beschäftigten und für die Frage nach der Schwarzarbeit für eine Firma wiederum besser als bei der Frage nach der Schwarzarbeit für eine Privatperson.

Von Interesse ist aber nicht nur der Vergleich zwischen den Erhebungsmethoden innerhalb unserer Studien, sondern auch der Vergleich mit den Prävalenzen anderer Studien, die in den letzten fünf Jahren in Deutschland durchgeführt wurden. Da diese

nicht immer nach Schwarzarbeit für eine Firma oder für eine Privatperson differenzieren, lässt sich dieser Vergleich nur für die Gesamtprävalenz anstellen. Zudem unterscheiden sich Populationen und Referenzzeiträume, so dass insbesondere zwei Studien mit Bezugszeiträumen von einem Jahr und der deutschen Wohnbevölkerung im Alter ab 15 Jahren im Eurobarometer (Williams 2009) respektive von 18 bis 74 Jahren (Feld & Larsen 2008) als Grundgesamtheit näherungsweise zum Vergleich herangezogen werden können.

Schätzt man aus den individuellen Angaben zur Schwarzarbeit für eine Privatperson und für eine Firma eine Gesamtprävalenz, so ergeben sich für die direkte Abfrage der RRT-Studie insgesamt 1,9 Prozent Schwarzarbeit in der Beschäftigtenstichprobe, während die Schätzung für die Leistungsempfänger bei insgesamt 3,8 Prozent liegt. In den jeweiligen Experimentalkonditionen der RRT-Studie ergeben sich folgende Untergrenzen: In der Beschäftigtenstichprobe liegt der Wert bei mindestens

1,2 Prozent in der Gruppe der Leistungsempfänger hingegen bei ca. 9,9 Prozent. In der direkten Abfrage der hier vorgestellten Bevölkerungsstichprobe der ICT-Studie ergeben sich insgesamt 4,2 Prozent Schwarzarbeit, ohne Differenzierung nach Auftraggeber, während dies in der ICT-Kondition mindestens 6,4 Prozent sind.⁸ Berücksichtigt man, dass beide Studien primär in den Monaten Oktober und November 2010 durchgeführt wurden, so dürfte der Jahreswert sogar noch etwas höher liegen als diese errechneten Anteile. Demgegenüber kommt der Eurobarometer auf einen geschätzten Anteilswert von 2,6 Prozent für 2007 (Williams 2009: 154) – ein Wert der sich in der Größenordnung in die entsprechenden Werte der direkten Konditionen beider von uns durchgeführten Studien einfügt. Die Jahresprävalenz der Studie von Feld und Larsen (2008) liegt dagegen bei 7,2 Prozent (im Jahr 2006).

6. Erklärung von Schwarzarbeit: Multivariate Analysen

In diesem Abschnitt soll mithilfe der in der RRT-Studie erhobenen Daten der Frage nachgegangen werden, welche individuellen Einflussfaktoren Schwarzarbeit begünstigen.⁹ Wir wollen zunächst kurz einen theoretischen Erklärungsrahmen hierfür vorstellen und die abgeleiteten Hypothesen anschließend mithilfe multipler Regressionsmodelle überprüfen. Wir leiten dabei unsere Hypothesen aus handlungstheoretischen Überlegungen (Becker 1968; Voss & Abraham 2000; Mehlkop & Becker 2004; Eifler 2009) sowie aus Theorien zur Steuerhinterziehung ab.

6.1 Theoretische Überlegungen und Operationalisierungen

Das Standardmodell der Steuervermeidung und -hinterziehung untersucht vor allem den Einfluss von Steuersätzen auf dieselbe (Allingham & Sandmo 1972). In ihrem Modell diskutieren die Autoren auch den Einfluss der (den Individuen als bekannt unterstellten) Wahrscheinlichkeit einer Steuerprüfung und das mit einer Entdeckung einhergehende Strafmaß. In den vergangenen Jahrzehnten haben sich die Modelle zur Steuerhinterziehung in verschiedene Richtungen weiterentwickelt. Begründet wird die Notwendigkeit dieser Weiterentwicklungen des Modells von Allingham und Sandmo damit, dass dieses einer empirischen Überprüfung nicht standhält.

So argumentiert Bordignon (1993: 345), dass es zum einen zu viel „Compliance“ gibt: Der hohe Anteil derjenigen Personen, die keine Steuern vermeiden oder hinterziehen kann nur unzureichend erklärt werden. Außerdem sind einige postulierte Zusammenhänge, z. B. dass die Höhe der hinterzogenen Steuern mit der Steuerquote negativ zusammenhängt (Yitzhaki 1974; Andreoni et al. 1998), nicht plausibel und empirisch widerlegt. So wird von einer – zumindest in gewissen Bereichen – positiven Funktion ausgegangen. Drittens zeigt sich, dass das Standardmodell weder mit Befragungsdaten noch mit Experimenten belegt werden kann.

Weiterentwicklungen des Standardmodells zur Erklärung von Steuerhinterziehung integrieren Eigenschaften des Steuersystems und der Strafregelungen (Yitzhaki 1974). Auch wurden realistischere Annahmen, wie beispielsweise unvollständige Informationen bezüglich Steuerprüfungen, eingeführt. Die Rolle von sozialen Normen sowie durch soziale Ächtung entstehende Kosten (Benjamini & Maital 1985) oder Konformitätsbestrebungen und Moral von Individuen (Gordon 1989; Myles & Naylor 1996), Fairnessabwägungen (Spicer & Lundstedt 1976; Bordignon 1993) sowie der Einfluss von Steuerberatern werden berücksichtigt. Dell'Anno (2009) beschreibt zwei Richtungen, in die sich die Forschung in diesem Bereich primär entwickelt hat: So integriert der Autor in seinem Ansatz die subjektive Entdeckungswahrscheinlichkeit und Erwartungsnutzentheorie aus der Verhaltensökonomie sowie die individuelle Steuermoral, definiert als intrinsische Motivation, Steuern zu bezahlen. Da bei (entdeckter) Steuerhinterziehung psychische Kosten durch Schamgefühle sowie Kosten durch Stigma und Verlust des Rufes entstehen, werden diese ebenfalls berücksichtigt (vgl. auch Gordon 1989).

⁸ Die Gesamtprävalenzen in den jeweiligen Kontrollgruppen ergeben sich durch Addition der Prävalenzen der beiden Items bereinigt um diejenigen Personen, die beide Arten von Schwarzarbeit ausüben (Vereinigungsmenge). Für die RRT-Kondition berechnet sich die Gesamtprävalenz gemäß den jeweiligen bedingten Wahrscheinlichkeiten (wobei die negative Schätzung des zweiten Items bei der Beschäftigtenstichprobe zur Berechnung bei null zensiert wurde). Für die ICT-Kondition wurde eine Untergrenze berichtet, die nur auf dem zweiten Item beruht, da hier die Berechnung der Gesamtprävalenz nicht möglich ist.

⁹ Aufgrund der zu geringen statistischen Power der ICT-Studie werden die nachfolgenden Analysen nur für die RRT-Studie durchgeführt.

Einen schematischen Überblick über die Entwicklung der verschiedenen Ansätze bieten Cowell (1990) und Andreoni et al. (1998).¹⁰

Basierend auf diesen Überlegungen nutzen wir klassische und neuere handlungstheoretische Argumentationsmuster zur Erklärung delinquenten Verhaltens (Becker 1968; Voss & Abraham 2000; Mehlkop & Becker 2004). Den Handlungsrahmen von Individuen bilden bei diesen Ansätzen neben der subjektiven Erfolgswahrscheinlichkeit und der damit zusammenhängenden Nutzenerwartung auch die antizipierten Kosten einer kriminellen Handlung, sollte diese entdeckt werden, und die damit zusammenhängende Entdeckungswahrscheinlichkeit. Zudem werden soziale Normen im persönlichen Umfeld und andere Restriktionen als relevant erachtet (Mehlkop 2011). Angewendet auf die individuelle Entscheidung einer Person, Schwarzarbeit anzubieten (Partizipationsentscheidung), sehen wir konkret die folgenden erklärenden Faktoren als handlungsrelevant an: Den *monetären Gewinn*, der durch die nicht deklarierte Arbeit erzielt wird, die *wahrgenommenen Kosten* von Schwarzarbeit, *Gelegenheitsstrukturen* sowie *Werte und Normen* im Hinblick auf die Akzeptanz von Schwarzarbeit.

Der *monetäre Gewinn*, den ein Individuum aus der Verrichtung von Schwarzarbeit erzielt, hängt vor allem vom erwarteten Mehreinkommen ab (Tanzi & Shome 1993: 811). Dieses Mehreinkommen kann in zweierlei Hinsicht erzielt werden. Zum einen durch Substitution gemeldeter Beschäftigung: Ein Schwarzarbeiter, der keine anrechenbaren Sozialleistungen wie Arbeitslosengeld oder Arbeitslosengeld II bezieht, spart dann Steuern und Sozialabgaben. Da die Sozialabgaben zumindest teilweise wieder der Person selbst zu Gute kommen (über die Höhe des Arbeitslosengelds oder die Rentenhöhe), wollen wir uns hier auf die Höhe des Grenzsteuersatzes konzentrieren. Für beschäftigte Personen variiert dieser, abhängig vom Einkommen, zwischen null (Einkommen unter 8.000 Euro pro Jahr) und 42 Prozent (ab 52.500 Euro pro Jahr). Noch deutlicher können Substitutionsgewinne ausfallen, wenn die Person Leistungen erhält, auf die

das Einkommen aus Erwerbstätigkeit anrechenbar ist. Für Empfänger von Arbeitslosengeld II liegt diese Transferenzugsrate zwischen 80 und 90 Prozent für jeden Euro, der über 100 Euro monatlich verdient wird.¹¹ Da der Grenzsteuersatz und die Transferenzugsrate in der vorliegenden Studie nicht detailliert gemessen werden, dient die Stichprobenzugehörigkeit in Kombination mit dem (Brutto-)Erwerbseinkommen als kruder Proxy für den Gewinn durch Substitution. Unsere Hypothese lautet, dass Personen im Grundsicherungsbezug und Personen mit hohem Erwerbseinkommen (Andreoni et al. 1998) *ceteris paribus* mehr schwarzarbeiten.

Ein Gewinn durch Schwarzarbeit lässt sich zum anderen in Form eines Zusatzeinkommens neben der regulären Erwerbstätigkeit erzielen. Dies ist dann der Fall, wenn ein gewünschter Beschäftigungsumfang am regulären Arbeitsmarkt nicht realisiert werden kann. So können bspw. Rahmentarifverträge individuelle Arbeitszeitwünsche beschränken, aber auch Beschäftigte unfreiwillig in Teilzeit beschäftigt werden. Als Indikator hierfür verwenden wir die gewünschte wöchentliche Arbeitszeit in Relation zur tatsächlich geleisteten wöchentlichen Arbeitszeit. Unsere Hypothese lautet, dass Personen, die ihre gewünschte Arbeitszeit nicht am regulären Arbeitsmarkt realisieren können, eher schwarzarbeiten als solche, die sie realisieren können.

Die *monetären Kosten* setzen sich aus dem wahrgenommenen Entdeckungsrisiko und der erwarteten Strafhöhe im Falle einer Entdeckung zusammen (vgl. Andreoni et al. 1998; Pedersen 2003; Feld & Larsen 2005). Formal gesehen ergibt sich der Erwartungswert der Kosten aus dem Produkt beider Merkmale. Je höher dieses Produkt für ein gegebenes Delikt ausfällt, desto geringer sollte die Schwarzarbeitsneigung einer Person sein. Gebildet wurde dieser Proxy aus zwei Angaben. Zum einen wurden die Befragten gebeten einzuschätzen, wie viele von 100 schwarzarbeitenden Personen von den Behörden entdeckt würden, zum anderen, welche Geldstrafe im Falle einer Entdeckung zu erwarten wäre (den Befragten wurde hier ein konkretes Fallbeispiel genannt).

Bisher lag der Fokus auf der Anreiz- und Motivationsstruktur des potenziellen Arbeitsanbieters. Die Partizipationsentscheidung hängt allerdings auch von der Nachfrage ab. *Gelegenheitsstrukturen* spie-

¹⁰ Grundsätzlich halten wir es in Anlehnung an Dell'Anno (2009: 993) für möglich, die (weiter)entwickelten Modelle zur Steuervermeidung und -hinterziehung auf Schwarzarbeit anzuwenden, da davon ausgegangen werden kann, dass es sich dabei um die gleichen Personen handelt, die auch in der Schattenwirtschaft tätig sind, also schwarzarbeiten. Weitere Ansätze finden sich u. a. bei Renooy 1990; Hessing et al. 1993; Lamnek et al. 2000; Mummert & Schneider 2001; Wenzel 2004; Feld & Larsen 2005; Schneider & Enste 2007: 54ff.

¹¹ Zur Zeit der Erhebung lag die Transferenzugsrate für jeden verdienten Euro zwischen 100 und 800 Euro im Monat bei 80 % und ab 800 Euro bis maximal 1.500 Euro bei 90 %.

len daher eine wesentliche Rolle für die Erklärung von Schwarzarbeit: Die Art der ausgeübten Tätigkeit, entweder im letzten oder im aktuellen Job, wird über zusammengefasste Kategorien der Erikson-Goldthorpe-Portocarero-Klassen (EGP) gemessen (Erikson et al. 1979). Frühere Studien zeigen etwa, dass selbständige Personen angesichts besserer Gelegenheiten mehr schwarzarbeiten (Andreoni et al. 1998; Feld & Larsen 2005; Williams 2009). Weiterhin zählt zur Gelegenheitsstruktur, wie gut ein Individuum in einer Region vernetzt bzw. integriert ist. Während gut vernetzte Personen auf der einen Seite mehr soziale Verpflichtungen und damit weniger Zeit haben schwarz zu arbeiten, bieten diese Netzwerke auf der anderen Seite auch mehr Informationen und Gelegenheiten zur Schwarzarbeit (Granovetter 1995[1974]; Wolff 1991; Williams 2010). Um diesen Effekt zu erfassen, wurde die Anzahl derjenigen Personen erhoben, die dem Befragten potenziell helfen könnten, einen Job zu finden. Als weiteren Indikator für die Vernetzung ziehen wir die aktive Vereinsmitgliedschaft heran. Unsere Hypothese lautet, dass Personen, die bezüglich der genannten Indikatoren besser vernetzt oder von der Art ihres Berufs her eigenständiger sind, mehr Gelegenheiten haben, schwarz zu arbeiten und daher auch häufiger schwarzarbeiten werden.

Schließlich sind normative Aspekte wie *informelle normative Erwartungen* oder *internalisierte Normen* für die Erklärung devianten Verhaltens relevant (Gordon 1989; Tanzi & Shome 1993; Cullis & Lewis 1997; Andreoni et al. 1998; Falk 2003; Feld & Larsen 2005; Cialdini 2007). Einerseits werden Personen, die selbst (oder deren soziales Umfeld) Schwarzarbeit ablehnend gegenüberstehen, gar nicht erst abwägen, ob ihr Nutzen durch Schwarzarbeit höher wäre als durch reguläre Arbeit; andererseits kann man die emotionalen Kosten eines Verstoßes gegen eigene normative Vorstellungen auch als Kostenkomponente in das Modell integrieren. So sprechen Andreoni et al. (1998: 846) beispielsweise vom „human need for consistent self-representation“ und, bei internalisierten Ehrlichkeitsnormen, von einer Tendenz zu normkonformem Verhalten. Auf der anderen Seite könnte die Wirkung dieser internalisierten moralischen Normen durch gegenläufige soziale Normen abgeschwächt bzw. sogar ins Gegenteil verkehrt werden, und zwar dann, wenn im sozialen Netzwerk des Befragten Schwarzarbeit weitgehend positiv bewertet, weit verbreitet und damit als „normal“ akzeptiert ist. Soziale Normen zu Schwarzarbeit, die in den eigenen Netzwerken vermittelt werden, werden anhand des von den Befragten geschätzten Anteils an

Schwarzarbeitern unter ihren Freunden und Bekannten sowie einen Index zur Akzeptanz von Schwarzarbeit gemessen. Dieser Index wurde mit sieben Einstellungsfragen erfasst, die sich unter anderem auf Steuermoral, wahrgenommene Regulierungsdichte oder Intoleranz gegenüber Schwarzarbeit beziehen. Die Angaben wurden jeweils auf vierstufigen Antwortskalen erhoben und, nachdem die Polung der einzelnen Items gleichgerichtet wurde, zu einem additiven Index¹² zusammengefasst. Hohe Skalenwerte drücken dabei eine positive Haltung zu Schwarzarbeit aus. Je größer die Ablehnung von Schwarzarbeit und je geringer der Anteil schwarzarbeitender Personen im eigenen Bekanntenkreis ist, desto geringer sollte die individuelle Wahrscheinlichkeit sein, selbst schwarz zu arbeiten.¹³

Zusammengefasst lassen sich die folgenden Hypothesen ableiten:

- 1) Personen im Grundsicherungsbezug, Personen mit einem hohen Erwerbseinkommen sowie Personen, die ihre gewünschte Arbeitszeit am regulären Arbeitsmarkt nicht realisieren können, haben eine höhere individuelle Neigung zur Schwarzarbeit (*Nutzen-Hypothese*).
- 2) Personen, die ein subjektiv höheres Entdeckungsrisiko und subjektiv höhere Strafkosten erwarten, haben eine geringere individuelle Wahrscheinlichkeit zur Schwarzarbeit (*Kosten-Hypothese*).
- 3) Personen, die sozial besser vernetzt sind (beispielsweise über Netzwerke in der Jobsuche oder aktive Vereinsmitgliedschaften) oder von der Art ihres Berufes her eigenständiger sind, haben eine höhere individuelle Wahrscheinlichkeit zur Schwarzarbeit (*Gelegenheitshypothese*).
- 4) Personen, die Schwarzarbeit ablehnen oder in deren wahrgenommenem Umfeld Schwarzarbeit seltener vorkommt, haben eine geringere individuelle Wahrscheinlichkeit zur Schwarzarbeit (*Normen-Hypothese*).

Außerdem kontrollieren wir in den nachfolgenden Analysen für demographische Merkmale wie Geschlecht, Alter, Ausbildung, Migrationshintergrund und Wohnort (Ost/West) (vgl. auch Wolff 1991;

¹² Ein Vergleich der Summenscores mit entsprechenden Faktorwerten (MLE) zeigt, dass diese hoch korreliert sind ($r = 0,95$; unabhängig davon ob imputiert wurde oder eine Analyse auf der Basis vollständiger Fälle durchgeführt wurde). Cronbach's Alpha beträgt 0,56 für die sieben Items.

¹³ Eine Übersicht der Operationalisierungen und der empirischen Verteilungen der Merkmale findet sich im Online-Anhang unter www.zfs-online.org.

Merz & Wolff 1993; Andreoni et al. 1998; Mummert & Schneider 2001; Schneider & Enste 2007; Williams 2009; Lago-Peñas & Lago-Peñas 2010).

6.2 Zentrale empirische Ergebnisse

Die nachfolgenden Analysen untersuchen mittels eines logistischen Regressionsansatzes (Jann 2011 mit modifizierter Likelihood-Funktion, vgl. Madala 1983) die postulierten Hypothesen zur Erklärung von Schwarzarbeit. Um die postulierten Zusammenhänge möglichst flexibel zu modellieren und eine sinnvolle statistische Analyse zu ermöglichen, wurden angesichts der relativ geringen Power des RRT-Schätzers überwiegend grobe Kategorisierungen (mit je 3 bis 4 disjunkten Kategorien aus den empirischen Terzilen bzw. Quartilen)¹⁴ zur Messung der erklärenden Variablen gebildet. Um die Power zu erhöhen, nutzen wir die kumulierten Daten der direkten Erhebung und der RRT auf der Grundlage beider Teilstichproben. Aufgrund der geringen Power des RRT-Schätzers wurde auf eine Gewichtung mit unterschiedlichen Auswahlwahrscheinlichkeiten (Designgewichten) verzichtet. Aus demselben Grund verzichten wir auf die Modellierung von Interaktionseffekten.

Um dem Problem von Ausfällen (Item-Nonresponse) zu begegnen, wurden fehlende Werte mittels eines Hot-Deck-Verfahrens (Mander & Clayton 1999) bzw. über eine Mittelwertimputation (im Falle des Merkmals Einstellungen Schwarzarbeit) ersetzt.¹⁵ Wir haben Sensitivitätsanalysen sowohl mit gewichteten als auch mit nicht imputierten Daten durchgeführt, deren Ergebnisse sich nicht grundlegend von den hier präsentierten unterscheiden. Allerdings sinkt aufgrund des Verlusts an statistischer Power der Anteil signifikanter Befunde.¹⁶

Tabelle 3 gibt die durchschnittlichen marginalen Effekte („average marginal effect“, AME) der logistischen Regressionsmodelle mit den jeweiligen Standardfehlern wieder (Long 1997; Allison 1999;

Bartus 2005; Mood 2009).¹⁷ Der AME bietet sich in der Interpretation an, da dieser über Modelle hinweg vergleichbar und intuitiv verständlich ist: Er ist ein Durchschnittseffekt der „unabhängigen Variablen als Mittelwert der marginalen Effekte über *alle* Beobachtungen“ (Best & Wolf 2012: 383). Steigt die unabhängige Variable x um eine Einheit, „so steigt die Wahrscheinlichkeit von $y = 1$ durchschnittlich um AME [Prozent]Punkte“ (Best & Wolf 2012: 384).

Nach der *Nutzen-Hypothese* erwarten wir eine höhere individuelle Schwarzarbeitsneigung bei Personen mit hohen Steuer- bzw. Transferentzugsraten und bei Personen, die ihre gewünschte Arbeitszeit nicht am Markt realisieren können. Personen, die ihre gewünschte Arbeitszeit nicht am Markt realisieren können, arbeiten zwar in beiden Modellen (Schwarzarbeit für eine Privatperson und Schwarzarbeit für eine Firma) häufiger schwarz, doch sind die Effekte klein und nicht signifikant.

Auch bezüglich der Entzugsrate (Steuersatz bzw. Transferentzug) sind die Vorzeichen überwiegend in Übereinstimmung mit der Hypothese. So weisen in beiden Modellen Grundsicherungsempfänger mit über 800 Euro Einkommen (90 %-Entzugsrate) jeweils die höchste und Grundsicherungsempfänger mit bis zu 800 Euro Einkommen (80 %-Entzugsrate) jeweils die zweithöchste Schwarzarbeitsneigung auf. Lediglich zwischen den Beschäftigten mit höheren (eher höhere Steuersätze) und geringeren (eher niedrigere Steuersätze) Einkommen sind die Ergebnisse zwischen den Modellen nicht konsistent.

Testet man alle Kategorien der kombinierten ALG-II-Bezugs- und Einkommensvariable gegeneinander¹⁸, so erweisen sich bei der Schwarzarbeit für eine Privatperson nur die Kontraste zwischen ALG-

¹⁴ Alternative Kategorisierungen führen zu den gleichen Schlussfolgerungen wie die berichteten.

¹⁵ Für die Variablen Entdeckungswahrscheinlichkeit und wahrgenommene Strafhöhe liegt das Ausmaß fehlender Werte bei 4,3 % respektive 10,0 %. Zur Prävalenz von Schwarzarbeit unter den Bekannten machten 5,7 % der Befragten keine Angabe. Auch die Anzahl derer, die einen Job vermitteln können, wurde in 4,7 % der Fälle nicht genannt. Die übrigen Variablen weisen maximal 2,7 % fehlende Werte auf.

¹⁶ Ergebnisse der Sensitivitätsanalysen sind auf Anfrage von den Autoren erhältlich.

¹⁷ Die AME wurden mit dem Stata-Befehl „margins“ geschätzt (Williams 2012). Hierbei werden die Standardfehler auf der Grundlage der Delta-Methode bestimmt (Oehlert 1992). Die empirischen Signifikanzniveaus (p-Werte) basieren jeweils auf zweiseitigen z-Tests.

¹⁸ Die Analysen enthalten eine große Zahl kategorialer Variablen. In Tabelle 3 sind für diese aus Platzgründen jeweils nur die AME und deren Standardfehler gegenüber einer Referenzkategorie ausgewiesen. Im Rahmen der Analysen wurden beide Modelle mit allen Kontrasten als Referenzkategorien berechnet. Die AME der übrigen Kontraste lassen sich durch Differenzbildung leicht berechnen. Sie werden im Text inklusive der zugehörigen p-Werte lediglich bei statistisch signifikanten Effekten ausgewiesen. Aufgrund von Rundungen kann der im Text berichtete AME auf der Nachkommastelle um 0,001 von den aus Tabelle 3 hervorgehenden Differenzen abweichen.

Tabelle 3 Logistische Regression für RRT-Designs (Stata-Befehl „rrlogit“; Jann 2011) mit Schwarzarbeit für eine Privatperson bzw. eine Firma als abhängige Variable (Average Marginal Effects; Standardfehler in Klammern)

AV: Schwarzarbeit für eine		Privatperson AME	Firma AME
Methodeneffekt			
Experimentalkondition (Ref. Direkt)	RRT	0,024* (0,011)	0,047** (0,012)
	Wechsler	-0,008 (0,017)	-0,007 (0,020)
Nutzen-Hypothese			
Nutzen von Schwarzarbeit (Ref. ALG-II & Eink. bis 800 €)	Beschäftigtenstichprobe & Einkommen bis 800 €	-0,030 (0,018)	-0,010 (0,015)
	Beschäftigtenstichprobe & mind. 800 € Einkommen	-0,025 (0,015)	-0,049* (0,023)
	ALG-II-Stichprobe & mind. 800 € Einkommen	0,010 (0,014)	0,010 (0,017)
Gewünschte Arbeitszeit: (Ref. bis 2 Stunden)	3 Stunden und mehr	0,008 (0,012)	0,004 (0,014)
Kosten-Hypothese			
Subj. Risikoeinschätzung: (Ref. 0 – 120 €)	geringes Risiko 121 € – 600 €	0,007 (0,011)	0,003 (0,013)
	hohes Risiko 601 € – 2.500 €	-0,002 (0,012)	-0,001 (0,012)
	sehr hohes Risiko > 2.500 €	-0,029+ (0,017)	-0,002 (0,012)
Gelegenheitshypothese			
Sozioökonomische Stellung: (Ref. Un/ Angelehrte)	TNZ (nie gearbeitet)	-0,016 (0,017)	-0,004 (0,013)
	Niedrige/ mittlere/ obere Ränge der Dienstklasse	-0,001 (0,015)	0,008 (0,016)
	Nicht manuelle Routinearbeit	-0,006 (0,014)	-0,006 (0,013)
	Selbständige	0,020 (0,022)	-0,020 (0,035)
	Meister/ Facharbeiter/ gelernte Arbeiter	0,025+ (0,013)	-0,014 (0,017)
Netzwerke (Jobsuche): (Ref. 0 Personen)	1 bis 4 Personen	0,031* (0,015)	0,005 (0,014)
	5 bis 10 Personen	0,038** (0,014)	0,017 (0,011)
	11 Personen und mehr	0,045** (0,016)	0,011 (0,014)
Akt. Vereinsmitgliedschaft: (Ref. Keine)	Mitglied in min. 1 Organisation	0,018+ (0,010)	0,009 (0,010)
Normen-Hypothese			
Prävalenz Bekanntenkreis: (Ref. Niemand)	1 bis 3 %	0,055* (0,021)	0,011 (0,016)
	4 bis 10 %	0,061** (0,020)	0,013 (0,016)
	11 bis 100 %	0,078** (0,020)	0,045** (0,014)

Tabelle 3 (Fortsetzung)

AV: Schwarzarbeit für eine		Privatperson	Firma
		AME	AME
Einstellungen Schwarzarbeit:	Eher contra... (15–16)	0,008 (0,020)	0,011 (0,014)
(Ref. Contra Schwarzarbeit 7–14)	Eher pro... (17–19)	0,022 (0,018)	0,001 (0,015)
	Pro... (20–28)	0,047** (0,017)	0,018 (0,013)
Kontrollvariablen			
Geschlecht: (Ref. Männlich)	Weiblich	-0,001 (0,010)	-0,019+ (0,011)
Alter: (Ref. 35 bis 49 Jahre)	< = 34 Jahre	0,012 (0,011)	0,030* (0,012)
	> = 50 Jahre	0,015 (0,013)	-0,039 (0,029)
Ausbildung: (Ref. Berufl. Ausbildung)	Schüler/ noch in bzw. keine Ausbildung	0,005 (0,012)	-0,002 (0,011)
	Mind. Fachhochschulabschluss	-0,064* (0,031)	-0,025 (0,030)
Migrationshintergrund: (Ref. Nein)	Migrationshintergrund	-0,006 (0,011)	-0,003 (0,010)
Region: (Ref. West)	Ost	-0,004 (0,011)	-0,001 (0,011)
Modellfit			
Pseudo R2		0,07	0,06
LR Chi2		125,16 (df=31)	90,01 (df=31)
AIC		1735,91	1597,92
BIC		1930,22	1792,24
N		3.204	3.205

+ p < 0.10, * p < 0.05, ** p < 0.01

II-Empfängern mit einem Einkommen ab 800 Euro und Beschäftigten mit einem Einkommen bis 800 Euro (AME = 0,040, p = 0,05) bzw. ab 800 Euro (AME = 0,034, p = 0,02) als signifikant.

Bei der Schwarzarbeit für eine Firma weisen dagegen Grundsicherungsempfänger (unabhängig vom Einkommen) nur gegenüber Beschäftigten mit einem Einkommen von über 800 Euro, eine signifikant erhöhte Schwarzarbeitswahrscheinlichkeit auf (bis 800 Euro: AME = 0,049, p = 0,03; über 800 Euro: AME = 0,059, p < 0,01). Beschäftigte mit einem Einkommen bis zu 800 Euro arbeiten zudem signifikant weniger schwarz als diejenigen mit einem Einkommen von mindestens 800 Euro (AME = 0,039, p = 0,07).

Insgesamt findet sich damit schwache Evidenz für die Hypothese, die aber nur im Vergleich zwischen

Grundsicherungsempfängern und Beschäftigten zu signifikanten Ergebnissen führt.

Die *Kosten-Hypothese* besagt, dass diejenigen Personen mit einer hohen Risikoeinschätzung bezüglich des wahrgenommenen Strafmaßes und der subjektiven Entdeckungswahrscheinlichkeit, seltener schwarzarbeiten als diejenigen mit einem geringen wahrgenommenen Risiko. Während sich diese Hypothese für das Item „Schwarzarbeit für eine Privatperson“ bestätigt – Personen, die das Strafrisiko sehr hoch einschätzen (> 2.500 Euro), arbeiten seltener schwarz als Personen mit sehr geringem (0–120 Euro: AME = -0,029, p = 0,09) oder geringem (121–600 Euro: AME = -0,036, p = 0,04) wahrgenommenem Risiko –, können wir diese Hypothese für das Modell „Schwarzarbeit für eine Firma“, bei der alle AMEs nahe bei null liegen, nicht bestätigen.

Gemäß der *Gelegenheitshypothese* würden wir erwarten, dass eigenständigere Berufe sowie eine stärkere Vernetzung Schwarzarbeit eher befördern. Unsere Ergebnisse zeigen, dass Meister bzw. Facharbeiter sowie gelernte Arbeiter häufiger für eine Privatperson schwarzarbeiten als an- und ungelernete Personen (AME = 0,025, $p = 0,06$). Dasselbe gilt gegenüber Personen, die noch nie erwerbstätig waren (AME = 0,042, $p = 0,03$) sowie gegenüber denjenigen, die einer nicht manuellen Routinetätigkeit nachgehen (AME = 0,031, $p = 0,04$). Aufgrund der geringen statistischen Power lassen sich trotz der teilweise stark von Null verschiedenen AMEs keine signifikant erhöhten Schwarzarbeitsquoten für Selbständige (AME = 0,020, $p = 0,37$) gegenüber Un-/Angelernten feststellen. Es zeigt sich in unseren Modellen jedoch, dass Personen mit größeren arbeitsmarktbezogenen Netzwerken, d. h. mit mindestens einer Person, die bei der Jobsuche behilflich sein kann (AME > = 0,031, $p < = 0,04$), oder Personen mit mindestens einer Vereinsmitgliedschaft (AME = 0,018; $p = 0,06$) häufiger für eine Privatperson schwarzarbeiten als Personen ohne entsprechende Netzwerke oder Vereinsmitgliedschaften. Im Firmenmodell lassen sich die Ergebnisse für die Gelegenheitshypothese nicht bestätigen: Keiner der erwarteten Effekte ist in diesem Modell statistisch signifikant.

Unsere Analysen bestätigen vor allem die postulierten Zusammenhänge im Rahmen der *Normen-Hypothese*: Schon eine geringe wahrgenommene Prävalenz von Schwarzarbeit im eigenen Bekanntenkreis (> = 1 %: AME > = 0,055, $p < = 0,01$) oder auch eine positive Einstellung zur Schwarzarbeit (AME = 0,047, $p < 0,01$) hängen positiv und statistisch signifikant mit der Schwarzarbeit für eine Privatperson zusammen. Letzteres gilt nicht nur in Relation zur jeweiligen Referenzkategorie des Modells: Auch der Kontrast zwischen Personen mit einer sehr positiven Einstellungen zu Schwarzarbeit (20–28) und den beiden Mittelkategorien ist signifikant (15–16: AME = 0,039, $p < 0,01$; 17–19: AME = 0,025, $p = 0,02$). Gleiches gilt für den Kontrast zwischen Personen, die einen hohen Anteil schwarzarbeitender Personen in ihrem persönlichen Netzwerk vermuten (> = 11 %), und denjenigen, die einen kleineren Anteil vermuten (4–10 %: AME = 0,024, $p = 0,07$).

Für das Item „Schwarzarbeit für eine Firma“, lässt sich die Normen-Hypothese zum Teil bestätigen: Personen, die einen sehr hohen Anteil an Schwarzarbeit im Freundeskreis vermuten (> = 11 %), arbeiten selbst auch signifikant häufiger schwarz als Personen, die keine Schwarzarbeit in ihrem sozialen

Umfeld berichten (AME = 0,045, $p < 0,01$). Dieser Effekt ist auch im Vergleich zu den beiden Mittelkategorien signifikant (1–3 %: AME = 0,034, $p = 0,01$; 4–10 %: AME = 0,031, $p = 0,01$).

Hinsichtlich der *Kontrollvariablen* zeigen unsere Modelle, dass sich Frauen und Männer im Personenmodell nicht unterscheiden, während Frauen im Firmenmodell auf dem 10 %-Niveau signifikant weniger schwarzarbeiten (AME = -0,019, $p = 0,07$). Für jüngere (bis 34) und ältere Personen (> = 50) lässt sich gegenüber den 35- bis 49-Jährigen kein Unterschied in Bezug auf die Häufigkeit von Schwarzarbeit im Personenmodell feststellen, während Personen bis 34 Jahre im Firmenmodell häufiger schwarzarbeiten als Personen ab 35 Jahren (< = 34: AME = 0,030, $p = 0,01$; > = 50: AME = 0,068, $p = 0,02$). Es zeigt sich außerdem, dass Personen mit (Fach-)Hochschulabschlüssen seltener für eine Privatperson schwarzarbeiten als Personen in der Referenzkategorie mit einer beruflichen Ausbildung (AME = -0,064, $p = 0,04$) oder auch als Schüler und Personen, die noch in Ausbildung sind (AME = -0,069, $p = 0,04$). Hinsichtlich der Schwarzarbeit für eine Firma zeigt sich bezüglich der Art des Ausbildungsabschlusses kein Unterschied. Auch die Kontrollvariablen Migrationshintergrund und Region zeigen in beiden Modellen keinen statistisch signifikanten Einfluss.

In beiden Modellen ist der *Methodeneffekt* in der erwarteten Richtung statistisch signifikant (AME = 0,024, $p = 0,03$; AME = 0,047, $p < 0,01$). Unter Kontrolle von erklärenden Drittvariablen und über beide Stichproben aggregiert berichten Personen in der RRT-Kondition bezüglich beider heikler Items signifikant häufiger schwarzarbeiten als Personen in der direkten Erhebungskondition. Dieser Effekt ist im Firmenmodell größer als im Personenmodell.

Zusammenfassend bestätigen sich unsere Hypothesen zur Erklärung von Schwarzarbeit eher für eine Privatperson als für eine Firma. Das ist zumindest zum Teil durch die geringere Prävalenz Letzterer zu erklären, die zu geringeren AMEs und geringerer statistischer Power führt. Die Vorzeichen der AMEs sind in beiden Modellen überwiegend identisch. Vor allem die postulierten Zusammenhänge im Rahmen der Gelegenheits- und der Normen-Hypothese hängen positiv und statistisch signifikant mit der Schwarzarbeit für eine Privatperson und auch mit der Schwarzarbeit für eine Firma zusammen.

7. Zusammenfassung und Diskussion

Die Studie hatte eine methodische und eine inhaltliche Zielsetzung. Zum einen ging es darum zu untersuchen, ob mit Befragungstechniken, die auf dem Prinzip der Anonymisierung von Antworten beruhen, validere Messungen von selbstberichteter Schwarzarbeit in Bevölkerungsbefragungen erzielt werden können und welche Probleme damit verbunden sind. Zum anderen ging es um die Erklärung von Schwarzarbeit durch individuelle Motivationslagen und situative Anreizstrukturen.

Als zentraler methodischer Befund lässt sich herausheben, dass sich weder die Item-Count- noch die Randomized-Response-Technik als der konventionellen direkten Befragung klar überlegen erwiesen haben. In insgesamt zwei von sechs Tests produzieren die RRT oder die ICT auf dem 5 %-Niveau signifikant höhere Prävalenzschätzungen als die direkte Befragung. Alle anderen Tests liefern keine statistisch signifikanten Unterschiede. Es ist dabei im Einklang mit dem Forschungsstand (Lensvelt-Mulders et al. 2005a), dass sich mit zunehmender Sensitivität des Items (Schwarzarbeit für eine Firma) bzw. in der Gruppe mit den schwerwiegenderen Straferwartungen (ALG-II-Empfänger) signifikante Effekte der Spezialtechniken zeigen.

Coutts und Jann (2011) geben außerdem zu bedenken, dass nicht nachgeprüft werden kann, ob Probanden überhaupt eine Münze zur Hand nehmen, d. h. ob sie gemäß den RRT-Instruktionen antworten. In der Tat legt der negative RRT-Schätzer den Schluss nahe, dass nicht immer gemäß den Regeln geantwortet wurde. Aus welchen Gründen die Regeln der RRT nicht befolgt wurden lässt sich mit der vorliegenden Studie nicht näher beleuchten. Zusammen mit den RRT-Wechsler*innen verweigerten insgesamt 369 Personen die Anwendung der RRT. Es haben also insgesamt ca. 15,8 Prozent der Befragten (vollständige und partielle Interviews betrachtet) der RRT-Kondition deren Anwendung abgelehnt. Diese Verweigerungsrate ist zusammen mit den deutlich erhöhten Kosten durch die längere Befragungszeit, der intensiveren Interviewerschulung und vor allem durch die bei gegebener statistischer Power deutlich höhere erforderliche Fallzahl ein starkes Argument gegen deren Einsatz.

Die Ergebnisse der ICT-Studie deuten zudem auf das Vorkommen sogenannter „ceiling effects“ (Glynn 2013) hin. Diese treten dann auf, wenn ein Befragter alle nicht sensitiven und das sensitive Item bejaht. In diesem Fall ist keine Anonymität mehr gewährleistet.

Neuere Studien argumentieren, dass die ICT der RRT überlegen ist, obwohl sie tendenziell statistisch noch ineffizienter ist (Corstange 2009; Holbrook & Krosnick 2010a, 2010b; Coutts & Jann 2011). Wo Befragte bei der RRT aus Unverständnis der Regeln oder Misstrauen nach dem „Haken“ suchen und deshalb (un)bewusst falsche Angaben machen, scheint die ICT in der Anwendung einfacher und überzeugender und somit praxistauglicher. Da die ICT im Gegensatz zur RRT keiner langwierigen Einführung und keines Zufallsgenerators bedarf, ferner allgemein mit geringeren kognitiven Belastungen für den Probanden einhergeht, wird ihr in der Fachliteratur generell größeres Potenzial für (die zumindest partielle) Lösung des Problems der sozialen Erwünschtheit zugeschrieben (Coutts & Jann 2011; Glynn 2013). In unserer Arbeit konnten wir diese Thesen allerdings empirisch nicht bestätigen.

Trotz dieser methodischen Schwierigkeiten und einer geringen statistischen Power des RRT-Schätzers, konnten die multivariaten Analysen einige der Hypothesen zu den individuellen Determinanten von Schwarzarbeit bestätigen. Als robuster Erklärungsfaktor für die inhaltlichen Modelle hat sich vor allem der wahrgenommene Anteil der Schwarzarbeit im Freundes- und Bekanntenkreis herausgestellt, der für beide abhängigen Variablen signifikant war (Normen-Hypothese). Daneben gibt es weitere Faktoren, die jeweils in einem der beiden Modelle signifikant sind und in dem anderen ebenfalls in die erwartete Richtung weisen: Neben dem Grundsicherungsbezug (der stellvertretend für einen hohen monetären Gewinn aus Schwarzarbeit steht), der Zahl der Freunde und Bekannten, die einem helfen können, einen Job zu finden, und der Mitgliedschaft in Vereinen und Organisationen, ist vor allem auch die Akzeptanz und positive Bewertung von Schwarzarbeit ein robuster Erklärungsfaktor der individuellen Neigung schwarzzuarbeiten (Nutzen-Hypothese und Gelegenheitshypothese). Zudem neigen jüngere Menschen bis einschließlich 34 Jahren eher zu Schwarzarbeit als Personen ab 35 Jahren. Die in anderen Studien berichteten Unterschiede zwischen Ost- und Westdeutschland (Mummert & Schneider 2001) können mit unserer Studie ebenso wenig repliziert werden wie Geschlechterunterschiede (Boockmann et al. 2010; Enste 2012).

Gemäß unseren empirischen Analysen kann also festgehalten werden, dass sich in der *Erklärung von Schwarzarbeit* vor allem die Normen-Hypothese bestätigt. Zudem zeigen die im Gesamtüberblick gemischten und wenig konsistenten methodischen

Befunde keine klare Überlegenheit der RRT oder der ICT gegenüber der direkten Befragung bei der *Messung von Schwarzarbeit*. Die Umsetzung dieser Verfahren in Bevölkerungsbefragungen geht zudem im Vergleich zu direkten Selbstauskünften mit erhöhten Kosten einher. So steigen die Kosten der Datenerhebung durch die Notwendigkeit einer größeren Stichprobe, einer komplexeren Interviewerschulung sowie einer längeren Befragungsdauer. Auf der Seite der Befragten steigen zudem die zeitlichen und kognitiven Belastungen beim Einsatz dieser Techniken. Es ist daher im Einzelfall genau abzuwägen, ob das erwartete Ausmaß der Reduktion sozial erwünschten Antwortverhaltens den erhöhten Datenerhebungsaufwand bei der Verwendung dieser Spezialtechniken rechtfertigt.

Literatur

- AAPOR – The American Association for Public Opinion Research, 2011: Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th Edition. AAPOR. Lanexa.
- Ahart, A.M. & P.R. Sackett, 2004: A New Method of Examining Relationships between Individual Difference Measures and Sensitive Behavior Criteria: Evaluating the Unmatched Count Technique. *Organizational Research Methods* 7: 101–114.
- Allingham, M.G. & A. Sandmo, 1972: Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 1: 323–338.
- Allison, P.D., 1999: Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods and Research* 28: 186–208.
- Andreoni, J., B. Erard & J. Feinstein, 1998: Tax Compliance. *Journal of Economic Literature* 36: 818–860.
- Barton, A.H., 1958: Asking the Embarrassing Question. *Public Opinion Quarterly* 22: 67–68.
- Bartus, T., 2005: Estimation of Marginal Effects Using Margeff. *Stata Journal* 5: 309–329.
- Becker, G.S., 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169–217.
- Benjamini, Y. & S. Maital, 1985: Optimal Tax Evasion and Optimal Tax Evasion Policy: Behavioral Aspects. S. 245–264 in: W. Gaertner & A. Wenig (Hrsg.), *The Economics of the Shadow Economy*. New York: Springer.
- Best, H. & C. Wolf, 2012: Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. *Kölnner Zeitschrift für Soziologie und Sozialpsychologie* 64: 377–395.
- Biemer, P., B.K. Jordan, M.L. Hubbard & D. Wright, 2005: A Test of the Item Count Methodology for Estimating Cocaine Use Prevalence. S. 149–174 in: J. Kennet & J. Gfroerer (Hrsg.), *Evaluating and Improving Methods Used in the National Survey on Drug Use and Health*. Rockville: Substance Abuse and Mental Health Service Administration.
- Boockmann, B., R. Döhrn, M. Groeneck & H. Verbeek, 2010: Abschätzung des Ausmaßes der Schwarzarbeit. Tübingen: Institut für Angewandte Wirtschaftsforschung.
- Bordignon, M., 1993: A Fairness Approach to Income Tax Evasion. *Journal of Public Economics* 52:345–62.
- Boruch, R.F., 1971: Assuring Confidentiality of Responses in Social Research: A Note on Strategies. *American Sociologist* 6: 308–311.
- Bradburn, N., S. Sudman & B. Wansink, 2004: *Asking Questions*. Revised Edition. San Francisco: Jossey-Bass.
- Breusch, T., 2005: Estimating the Underground Economy using MIMIC Models. Working Paper. National University of Australia. Canberra. <http://steconomice.uoradea.ro/leonardo3/pdf/ESTIMATING%20THE%20UNDERGROUND%20ECONOMY%20USING%20MIMIC%20MODELS.pdf> (Zugriff: 22.04.2013).
- Buehn, A., A. Karmann & F. Schneider, 2009: Shadow Economy and Do-it-yourself Activities: The German Case. *Journal of Institutional and Theoretical Economics* 165: 701–722.
- Cialdini, R.B., 2007: Descriptive Social Norms as Underappreciated Sources of Social Control. *Psychometrika* 72: 263–268.
- Cohen, J., 1988: *Statistical Power Analysis for the Behavioral Sciences*. Second edition. Hillsdale: Erlbaum.
- Corstange, D., 2009: Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis* 17: 45–63.
- Coutts, E. & B. Jann, 2011: Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). *Sociological Methods & Research* 40: 169–193.
- Coutts, E., B. Jann, I. Krumpal & A.-F. Näher, 2011: Plagiarism in Student Papers: Prevalence Estimates Using Special Techniques for Sensitive Questions. *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)* 231: 749–760.
- Cowell, F.A., 1990: *Cheating the Government: The Economics of Evasion*. London: MIT Press.
- Cullis, J.G. & A. Lewis, 1997: Why People Pay Taxes. From a Conventional Economic Model to a Model of Social Convention. *Journal of Economic Psychology* 18: 305–321.
- Dalton, D.R., J.C. Wimbush & C.M. Daily, 1994: Using the Unmatched Count Technique (UCT) to Estimate Base Rates for Sensitive Behavior. *Personnel Psychology* 47: 817–828.
- Deffaa, W., 1982: Anonymisierte Befragungen mit zufallsverschlüsselten Antworten. Die Randomized-Response-Technik (RRT): Methodische Grundlagen, Modelle und Anwendungen. Reihe: Hohenheimer volkswirtschaftliche Schriften, Band 1. Frankfurt a.M.: Lang.
- Dell’Anno, R., 2009: Tax Evasion, Tax Morale and Policy Maker’s Effectiveness. *Journal of Socio-Economics* 38: 988–997.
- Droitcour, J., R.A. Caspar, M.L. Hubbard, T.L. Parsley, W. Visscher & T.M. Ezzati, 1991: The Item Count

- Technique as a Method of Indirect Questioning: A Review of its Development and a Case Study Application. S. 185–210 in: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S. Sudman (Hrsg.), *Measurement Errors in Surveys*. New York: Wiley.
- EC – European Commission, 2007: Special Eurobarometer 284/Wave 67.3: Undeclared Work in the European Union. Report. Brussels. http://ec.europa.eu/public_opinion/archives/ebs/ebs_284_en.pdf (Zugriff: 22.04.2013).
- Eifler, S., 2009: *Kriminalität im Alltag: Eine handlungstheoretische Analyse von Gelegenheiten*. Wiesbaden: VS.
- Enste, D.H., 2012: Schwarzarbeit und Schattenwirtschaft. *Wirtschaftsdienst* 92(2): 136–138.
- Enste, D.H. & F. Schneider, 2006: Welchen Umfang haben Schattenwirtschaft und Schwarzarbeit? Ein Versuch zur Lösung des Rätsels. *Wirtschaftsdienst* 86(3): 85–198.
- Erikson, R., J.H. Goldthorpe & L. Portocarero, 1979: Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. *British Journal of Sociology* 30: 341–415.
- Evers, H.D., 1987: Subsistenzproduktion. Markt und Staat. Der sog. Bielefelder Verflechtungsansatz. *Geographische Rundschau* 39: 136–140.
- Falk, A., 2003: Homo Oeconomicus versus Homo Reciprocans: Ansätze für ein neues Wirtschaftspolitisches Leitbild? *Perspektiven der Wirtschaftspolitik* 4: 141–172.
- Feige, E. L., 1990: Defining and Estimating Underground and Informal Economies: The New Institutional Economics Approach. *World Development* 18: 989–1002.
- Feld, L.P. & C. Larsen, 2005: Black Activities in Germany in 2001 and in 2004. A Comparison Based on Survey Data. Kopenhagen: Rockwool Foundation.
- Feld, L.P. & C. Larsen, 2008: „Black“ Activities Low in Germany in 2006. Kopenhagen: Rockwool Foundation.
- Feld, L.P. & F. Schneider, 2010: Survey on the Shadow Economy and Undeclared Earnings in OECD Countries. *German Economic Review* 11: 109–149.
- Feld, L.P., A.J. Schmidt & F. Schneider, 2007: Tax Evasion, Black Activities and Deterrence in Germany: An Institutional and Empirical Perspective. Discussion Paper, presented at the Public Choice Meeting in Amsterdam, 29. March – 1. April 2007. <http://www.econ.jku.at/members/schneider/files/publications/deterrence.pdf> (Zugriff: 22.04.2013).
- Fox, J.A. & P.E. Tracy, 1986: *Randomized Response. A Method for Sensitive Surveys*. London: Sage.
- Glynn, A.N., 2013: What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment. *Public Opinion Quarterly* 77: 159–172.
- Gordon, J., 1989: Individual Morality and Reputation Costs as Deterrents to Tax Evasion. *European Economic Review* 33: 797–805.
- Granovetter, M., 1995[1974]: *Getting a Job. A Study of Contacts and Careers*. Second edition. Chicago Press: Chicago.
- Greenberg, B.G., A.-L.A. Abul-El, W.R. Simmons & D.G. Horvitz, 1969: The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association* 64 (326): 520–539.
- Greenberg, B.G., R.R. Kuebler, Jr., J.R. Abernathy & D.G. Horvitz, 1971: Application of the Randomized Response Technique in Obtaining Quantitative Data. *Journal of the American Statistical Association* 66 (334): 243–250.
- Häder, S. & S. Gabler, 1998: Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. S. 69–88 in: S. Gabler, S. Häder & J.H.P. Hoffmeyer-Zlotnik (Hrsg.), *Telefonstichproben in Deutschland*. Opladen: Westdeutscher Verlag.
- Hausman, J., 2001: Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives* 15(4): 57–67.
- Hendrickx, J., 2002: ISKO: Stata module to recode 4 digit ISCO-88 occupational codes. Statistical Software Components S425802, Boston College Department of Economics (revised 20 Oct 2004).
- Hessing, D.J., H. Elffers, H.S.J. Robben & P. Webley, 1993: Needy or Greedy? The Social Psychology of Individuals Who Fraudulently Claim Unemployment Benefits. *Journal of Applied Social Psychology* 23(3): 226–243.
- Holbrook, A.L. & J.A. Krosnick, 2010a: Social Desirability Bias in Voter Turnout Reports. Tests Using the Item Count Technique. *Public Opinion Quarterly* 74:37–67.
- Holbrook, A.L. & J.A. Krosnick, 2010b: Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity. *Public Opinion Quarterly* 74(2): 328–343.
- Horvitz, D.G., B.V. Shah & W.R. Simmons, 1967: The Unrelated Question Randomized Response Model. *Proceedings of the Social Statistics Section. American Statistical Association*. 65–72.
- IAW – Institut für Angewandte Wirtschaftsforschung e.V., 2013: Schattenwirtschaftsprognose 2013: Relativ günstige Wirtschaftsentwicklung und Entlastungen bei der Rentenversicherung führen zu weniger Schattenwirtschaft. Pressemitteilung. http://www.iaw.edu/w/IAWPDF.php?id=968&name=iaw_pm_01_2013_Schattenwirtschaftsprognose_2013.pdf (Zugriff: 22.04.2013).
- Jann, B., 2011: RRLOGIT: Stata Module to Estimate Logistic Regression for Randomized Response Data. Boston: Boston College Department of Economics.
- Koch, W.A.S., 2005: *Das Schwarzarbeit-Änigma*. *Wirtschaftsdienst* 85: 715–723.
- Krumpal, I., 2010: Sensitive Questions and Measurement Error: Using the Randomized Response Technique to Reduce Social Desirability Bias in CATI Surveys. Dissertation. Universität Leipzig.
- 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: 1387–1403.
- Krumpal, I., 2013: Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity* 47: 2025–2047.

- Krumpal, I. & A.-F. Näher, 2012: Entstehungsbedingungen sozial erwünschten Antwortverhaltens: Eine experimentelle Studie zum Einfluss des Wording und des Kontexts bei unangenehmen Fragen. *Soziale Welt* 63: 65–89.
- Kuk, A.Y.C., 1990: Asking Sensitive Questions Indirectly. *Biometrika* 77: 436–438.
- Kuklinski, J.H., P.M. Sniderman, K. Knight, T. Piazza, P.E. Tetlock, G.R. Lawrence & B. Mellers, 1997: Racial Prejudice and Attitudes toward Affirmative Action. *American Journal of Political Science* 41: 402–419.
- LaBrie, J.W. & M. Earleywine, 2000: Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique. *Journal of Sex Research* 37: 321–326.
- Lago-Peñas, I. & S. Lago-Peñas, 2010: The Determinants of Tax Morale in Comparative Perspective: Evidence from European Countries. *European Journal of Political Economy* 26: 441–453.
- Lara, D., S.G. García, C. Ellertson, C. Camlin & J. Suárez, 2006: The Measure of Induced Abortion Levels in Mexico Using Random Response Technique. *Sociological Methods and Research* 35: 279–301.
- Lara, D., J. Strickler, C.D. Olavarrieta & C. Ellertson, 2004: Measuring Induced Abortion in Mexico. *Sociological Methods & Research* 32: 529–558.
- Lee, R.M., 1993: *Doing Research on Sensitive Topics*. London: Sage.
- Lensvelt-Mulders, G.J.L.M. & H.R. Boeije, 2007: Evaluating Compliance with a Computer Assisted Randomized Response Technique: A Qualitative Study into the Origins of Lying and Cheating. *Computers in Human Behavior* 23: 591–608.
- Lensvelt-Mulders, G.J.L.M., J.J. Hox, P.G.M. van der Heijden & C.J.M. Maas, 2005a: Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. *Sociological Methods and Research* 33: 319–348.
- Lensvelt-Mulders, G.J.L.M., J.J. Hox & P.G.M. van der Heijden, 2005b: How to Improve the Efficiency of Randomized Response Design. *Quality & Quantity* 39: 253–265.
- Long, S.J., 1997: *Regression Models for Categorical and Limited Dependent Variables*. London: Sage.
- Maddala, G.S., 1983: *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.
- Mander, A. & D. Clayton, 1999: HOTDECK: Stata Module to Impute Missing Values Using the Hotdeck Method. Statistical Software Components S366901. Boston College Department of Economics, revised 02 Sep 2007.
- Mangat, N.S., 1994: An Improved Randomized Response Strategy. *Journal of the Royal Statistical Society (Series B)* 56: 93–95.
- Mangat, N.S. & R. Singh, 1990: An Alternative Randomized Response Procedure. *Biometrika* 77: 439–442.
- Mehlkop, G., 2011: Kriminalität als rationale Wahlhandlung: Eine Erweiterung des Modells der subjektiven Werterwartung und dessen empirische Überprüfung. Wiesbaden: VS.
- Mehlkop, G. & R. Becker, 2004: Soziale Schichtung und Delinquenz. Eine empirische Anwendung eines Rational-Choice-Ansatzes mit Hilfe von Querschnittsdaten des ALLBUS 1990 und 2000. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56: 95–126.
- Merz, J. & K.G. Wolff, 1993: The Shadow Economy: Illicit Work and Household Production: A Microanalysis of West Germany. *Review of Income and Wealth* 39: 177–194.
- Mood, C., 2010: Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26: 67–82.
- Moors, J.J.A., 1971: Optimization of the Unrelated Randomized Response Model. *Journal of the American Statistical Association* 66: 627–629.
- Mummert, A. & F. Schneider, 2001: The German Shadow Economy: Parted in a United Germany? *FinanzArchiv* 58: 286–316.
- Myles, G.D. & R.A. Naylor, 1996: A Model of Tax Evasion with Group Conformity and Social Customs. *European Journal of Political Economy* 12: 49–66.
- OECD – Organisation for Economic Co-operation and Development, 2002: *Measuring the Non-Observed Economy. A Handbook*. Paris: OECD.
- Oehlert, G.W., 1992: A Note on the Delta Method. *American Statistician* 46: 27–29.
- Pedersen, S., 2003: *The Shadow Economy in Germany, Great Britain and Scandinavia. A Measurement Based on Questionnaire Surveys*. Study No. 10. Copenhagen: Rockwool Foundation.
- Pfau-Effinger, B., 2009: Varieties of Undeclared Work in European Societies. *British Journal of Industrial Relations* 47: 79–99.
- Pickard, M. & J. Sardà, 2011: The Size of the Underground Economy in Germany: A Correction of the Record and New Evidence from the Modified-Cash-Deposit-Ratio Approach. *European Journal of Law and Economics* 32: 143–163.
- Raghavarao, D. & W.T. Federer, 1979: Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society (Series B)* 41: 40–45.
- Renooy, P., 1990: *The Informal Economy: Meaning, Measurement and Social Significance*. Netherlands Geographical Studies No. 115. Amsterdam: National Geographical Studies Association.
- Renooy, P., S. Ivarsson, O. van der Wusten-Gritsai & R. Meijer, 2004: *Undeclared Work in an Enlarged Union: An Analysis of Shadow Work – An In-Depth Study of Specific Items*. Brüssel: TNS.
- Schneider, F., 2003: The Shadow Economy. S. 286–296 in: C.K. Rowley & F. Schneider (Hrsg.), *Encyclopedia of Public Choice*. Dordrecht: Kluwer.
- Schneider, F., 2008: The Shadow Economy in Germany: A Blessing or a Curse for the Official Economy? *Economic Analysis & Policy* 38: 89–111.
- Schneider, F. & D.H. Enste, 2000: Shadow Economies: Size, Causes and Consequences. *Journal of Economic Literature* 38: 77–114.

- Schneider, F. & D.H. Enste, 2007: *The Shadow Economy. An International Survey*. New York: Cambridge University Press.
- Schneider, F., J. Volkert & S. Caspar, 2002: Schattenwirtschaft und Schwarzarbeit: Beliebt bei Vielen – Problem für Alle. Baden-Baden: Nomos.
- Smith, L.L., W.T. Federer & D. Raghavarao, 1974: A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions. *Proceedings of the American Statistical Association (Social Statistics Section)*: 447–452.
- Spicer, M.W. & S.B. Lundstedt, 1976: Understanding Tax Evasion. *Public Finance* 31: 295–305.
- Tanzi, V. & P. Shome, 1993: A Primer on Tax Evasion. *International Monetary Fund, Staff Papers* 40: 807–828.
- Thießen, U., 2011: Schattenwirtschaft: Vorsicht vor hohen Makroschätzungen. *Wirtschaftsdienst* 3: 194–201.
- Thomas, J., 1999: Quantifying the Black Economy: “Measurement Without Theory” Yet Again? *Economic Journal* 109(456): 381–389.
- Tourangeau, R. & T.W. Smith, 1996: Asking Sensitive Questions. The Impact of Data Collection Mode, Question Format and Question Context. *Public Opinion Quarterly* 60: 275–304.
- Tourangeau, R. & T. Yan, 2007: Sensitive Questions in Surveys. *Psychological Bulletin* 133(5): 859–883.
- Tsuchiya, T. & Y. Hirai, 2010: Elaborate Item Count Questioning: Why Do People Underreport in Item Count Responses? *Survey Research Methods* 4: 139–149.
- Tsuchiya, T., Y. Hirai & S. Ono, 2007: A Study of the Properties of the Item Count Technique. *Public Opinion Quarterly* 71: 253–272.
- Umesh, U. & R. Peterson, 1991: A Critical Evaluation of the Randomized Response Method. *Sociological Methods and Research* 20: 104–138.
- Van der Heijden, P.G.M., G. van Gils, J. Bouts & J.J. Hox, 2000: A Comparison of Randomized Response, Computer-Assisted Self-Interview, and Face-to-Face Direct Questioning: Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit. *Sociological Methods and Research* 28: 505–537.
- Voss, T. & M. Abraham, 2000: Rational Choice Theory in Sociology: A Survey. S. 50–82 in: S.R. Quah & A. Sales (Hrsg.), *The International Handbook of Sociology*. London: Sage.
- Warner, S.L., 1965: Randomized-Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63–69.
- Weissman A.N., R.A. Steer & D.S. Lipton, 1986: Estimating Illicit Drug Use Through Telephone Interviews and the Randomized Response Technique. *Drug and Alcohol Dependence* 18: 225–233.
- Wenzel, M., 2004: The Social Side of Sanctions: Personal and Social Norms as Moderators of Deterrence. *Law and Human Behavior* 28: 547–567.
- Williams, C.C., 2009: Formal and Informal Employment in Europe: Beyond Dualistic Representations. *European Urban and Regional Studies* 16: 147–159.
- Williams, C.C., 2010: Rethorizing Participation in the Underground Economy. *Labor Studies Journal* 35: 246–267.
- Williams, R., 2012: Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects. *Stata Journal* 12: 308–331.
- Wimbush, J.C. & D.R. Dalton, 1997: Base Rate for Employee Theft: Convergence of Multiple Methods. *Journal of Applied Psychology* 82: 756–763.
- Wolff, K., 1991: Schwarzarbeit in der Bundesrepublik Deutschland. Eine mikroanalytische Untersuchung. Frankfurt a.M.: Campus.
- Yitzhaki, S., 1974: A Note on Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 3: 201–202.
- Zoll, 2011: Ahndung von Ordnungswidrigkeiten und Verfolgung von Straftaten. <http://www.zoll.de/DE/Fachthemen/Arbeit/Bekaempfung-der-Schwarzarbeit-und-illegalen-Beschaeftigung/Ahndung-von-Ordnungswidrigkeiten-und-Verfolgung-von-Straftaten/ahndung-von-ordnungswidrigkeiten-und-verfolgung-von-straftaten.html> (Zugriff: 22.04.2013).

Autorenvorstellung

Antje Kirchner, geb. 1982 in Erfurt. Studium der Politik- und Verwaltungswissenschaften in Konstanz von 2002–2008. Promotionsstudentin am Institut für Statistik der Ludwig-Maximilians-Universität München. Seit 2010 wissenschaftliche Mitarbeiterin im Forschungsbereich „Panel Arbeitsmarkt und soziale Sicherung“ und im Kompetenzzentrum Empirische Methoden am Institut für Arbeitsmarkt- und Berufsforschung.

Forschungsschwerpunkte: Survey Methodology, Soziale Normen und soziale Erwünschtheit, Arbeitsmarktsoziologie, Vergleichende Politikwissenschaft.

Wichtigste Publikationen: *Social Capital and Unemployment: A Macro-Quantitative Analysis of the European Regions* (mit M. Freitag), *Political Studies* 59, 2011; *Political Parties and Higher Education Spending: Who Favours Redistribution?* (mit C. Rauh & R. Kappe), *West European Politics* 34, 2011; *Crafting Tolerance: The Role of Political Institutions in a Comparative Perspective* (mit M. Freitag & C. Rapp), *European Political Science Review* 3, 2011.

Ivar Krumpal, geb. 1975 in Bratislava. Studium der Verwaltungswissenschaft an der Universität Konstanz von 1997–2004. Promotion an der Fakultät für Sozialwissenschaften und Philosophie der Universität Leipzig in 2010. Seit 2005 wissenschaftlicher Mitarbeiter am Institut für Soziologie der Universität Leipzig (Lehrstuhl für Theorie und Theoriesgeschichte).

Forschungsschwerpunkte: Soziale Normen und soziale Erwünschtheit, Arbeitsmarktsoziologie, Antisemitismus, politische Wahlen, Methoden der empirischen Sozialforschung, Survey Methodology.

Wichtigste Publikationen: Determinants of Social Desirability Bias in Sensitive Surveys: A Literature Review. *Quality & Quantity* 47, 2013; Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning. *Social Science Research* 41, 2012; Asking Sensitive Questions Using the Cross-wise Model: An Experimental Survey Measuring Plagiarism (mit B. Jann & J. Jerke), *Public Opinion Quarterly* 76, 2012.

Mark Trappmann, geb. 1970 in Duisburg. Studium der Mathematik und Sozialwissenschaften in Duisburg und Groningen. Promotion an der Universität Duisburg-Essen. Von 2004–2006 Assistent an der Universität Konstanz. Seit 2007 Leiter des Forschungsbereichs „Panel Arbeitsmarkt und soziale Sicherung“ am Institut für Arbeitsmarkt- und Berufsforschung. Seit 2012 Professor für Soziologie, insbes. Survey-Methodologie an der Otto-Friedrich-Universität Bamberg.

Forschungsschwerpunkte: Survey Methodology, Soziale Netzwerke, Arbeitsmarktsoziologie.

Wichtigste Publikationen: Strukturanalyse sozialer Netzwerke: Konzepte, Modelle, Methoden (mit H.J. Hummell & W. Sodeur), Wiesbaden 2011; Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data (mit F. Kreuter & G. Müller). *Public Opinion Quarterly* 74, 2010.

Hagen von Hermanni, geb. 1985 in Hannover. Studium der Soziologie und Betriebswirtschaftslehre an der Universität Leipzig. Seit März 2012 wissenschaftliche Hilfskraft und seit Oktober 2012 wissenschaftlicher Mitarbeiter an der Erziehungswissenschaftlichen Fakultät der Universität Leipzig. Seit November 2012 zudem wissenschaftlicher Mitarbeiter am Zentrum für Lehrerbildung und Schulforschung der Universität Leipzig.

Forschungsschwerpunkte: Schul- und Unterrichtsforschung, neue Medien im Unterricht, nachhaltige Entwicklung.