

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel study at DIW Berlin

981-2018

Machbarkeitsstudie zur Verbesserung der Forschungsdateninfrastruktur im Bereich Hochvermögender mit dem Sozio-oekonomischen Panel (SOEP)

Carsten Schröder, Charlotte Bartels, Markus M. Grabka, Martin Kroh, Rainer Siegers

SOEPPapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at
<http://www.diw.de/soeppapers>

Editors:

Jan **Goebel** (Spatial Economics)
Stefan **Liebig** (Sociology)
David **Richter** (Psychology)
Carsten **Schröder** (Public Economics)
Jürgen **Schupp** (Sociology)

Conchita **D'Ambrosio** (Public Economics, DIW Research Fellow)
Denis **Gerstorf** (Psychology, DIW Research Fellow)
Elke **Holst** (Gender Studies, DIW Research Director)
Martin **Kroh** (Political Science, Survey Methodology)
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Fellow)
Thomas **Siedler** (Empirical Economics, DIW Research Fellow)
C. Katharina **Spieß** (Education and Family Economics)
Gert G. **Wagner** (Social Sciences)

ISSN: 1864-6689 (online)

German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: soeppapers@diw.de



**Machbarkeitsstudie zur Verbesserung der
Forschungsdateninfrastruktur im Bereich
Hochvermögender mit dem Sozio–oekonomischen
Panel (SOEP)**

Autoren: Carsten Schröder, Charlotte Bartels, Markus M. Grabka,
Martin Kroh, Rainer Siegers

Projektleitung: Carsten Schröder

Antragsteller: Charlotte Bartels, Jan Goebel, Markus M. Grabka,
Martin Kroh, Daniel Schnitzlein, Carsten Schröder, Jürgen Schupp und
Rainer Siegers

Mitarbeit: Paul Brockmann, Konstantin Goebler, Christoph Halbmeier,
Heike Nachtigall, Fabian Nemecek, Christopher Prömel, Axel Ramstein
und Thomas Rieger

DIW Berlin / SOEP, April 2018

Executive Summary

Das Ausmaß der Vermögenskonzentration in Deutschland lässt sich aufgrund der stark eingeschränkten Datenlage zu hohen Vermögen kaum abschätzen. Dementsprechend sieht auch der fünfte Armuts- und Reichtumsbericht der Bundesregierung weiteren Bedarf, die Datenlage in diesem Bereich zu verbessern (Bundesregierung, 2017). Wir haben eine Machbarkeitsstudie durchgeführt, in der ein innovatives Konzept entwickelt, implementiert und im Feld getestet wurde, um Personen mit hohen Vermögen in Deutschland (Top-Vermögende) gezielt zu befragen und in das Sozio-ökonomische Panel (SOEP) zu integrieren.

Damit sollen künftig repräsentative Aussagen für „Top-Vermögende“ ermöglicht werden. *Unter Top-Vermögende sind Personen zu verstehen, deren Vermögen im obersten Bereich existierender Bevölkerungsbefragungen liegen oder darüber (mind. hohe einstellige individuelle Nettovermögen).*

Die Ergebnisse dieser Machbarkeitsstudie sind Gegenstand des vorliegenden Berichts. Dieser konzentriert sich auf das bei Verteilungsanalysen häufig verwendete Aggregat des Nettovermögens, also die Summe aus Geld- und Sachvermögen abzüglich Schulden.

Es zeigt sich, dass eine Umsetzung des vorgeschlagenen Konzepts im Rahmen einer Hauptuntersuchung einen nachhaltigen Beitrag zur Verbesserung der Forschungsdateninfrastruktur im Bereich Top-Vermögender liefern sollte. Dies wäre ein wichtiger Beitrag zur Verbesserung der Datenlage in Deutschland in Hinblick auf die künftige Armuts- und Reichtumsberichterstattung sowie für die Verteilungsforschung insgesamt. Das Konzept lässt

sich in analoger Weise auch in anderen Ländern umsetzen. Somit könnte eine Hauptuntersuchung als internationales Vorreitermodell dienen.

Die grundlegende Idee unseres Konzepts basiert auf der empirischen Regelmäßigkeit, dass Top-Vermögende zumindest einen Teil ihres Vermögens in Form von Beteiligungen an Unternehmen halten.¹ Unternehmen sind wiederum verpflichtet, Informationen über Eigentümerstrukturen zu veröffentlichen. Der Dienstleister Bureau van Dijk (BvD) hat diese weltweit veröffentlichten Daten in einer Unternehmensdatenbank gesammelt, in der detaillierte Daten über 270 Mio. Unternehmen enthalten sind. Anhand dieser Datenbank identifizieren wir Personen mit Wohnsitz in Deutschland, die weltweit Beteiligungen halten. Aus dieser Personendatenbank haben wir eine über die Höhe ihrer Beteiligungen geschichtete und regional geclusterte Zufallsstichprobe von 2.000 Personen gezogen. Die Erwartung ist, dass mit der Höhe der Beteiligung, also mit dem Stratum, die Wahrscheinlichkeit steigt, zu den Top-Vermögenden zu gehören. Aus dieser stratifizierten Zufallsstichprobe wurden 124 Personen mit Standard-Instrumenten des SOEP erfolgreich befragt – auch zu ihren Vermögen.

Die zentralen Ergebnisse der Machbarkeitsstudie sind wie folgt:

1. Aus der BvD-Datenbank lassen sich ca. 1,5 Mio. Personen mit Wohnsitz in Deutschland ziehen, die „nennenswerte“ Unternehmensbeteiligungen im In- und Ausland halten – d.h. solche im Ausmaß von min-

¹Unsere Berechnungen unter Verwendung des Surveys Private Haushalte und Finanzen (PHF) zeigen, dass der Anteil der Haushalte, die Unternehmensbeteiligungen halten, mit der Höhe des Haushaltsnettovermögens positiv korreliert. Insbesondere zeigt sich ein dynamischer Anstieg des Anteils über die höchsten Perzentile der in Surveys beobachteten Verteilung der Haushaltsnettovermögen. So steigt der Anteil der Haushalte mit Beteiligungen von ca. 20 Prozent in den niedrigen auf mehr als 80 Prozent im obersten Perzentil.

destens 0,01 Prozent. *Die hohe Fallzahl erlaubt auch eine umfangreiche Implementierung des Konzepts im Rahmen einer Hauptuntersuchung.*

2. Die reichsten Deutschen laut Manager-Magazin sind auch in der BvD-Datenbank enthalten. *Dies lässt erwarten, dass sich mit dem Konzept tatsächlich Vermögende bis in den obersten Bereich der deutschen Vermögensverteilung abbilden lassen.*
3. Der Wert der individuellen Unternehmensbeteiligungen ist in der BvD-Datenbank nicht enthalten, wohl aber die Höhe der Beteiligung in Prozent und Finanzkennziffern zu den Unternehmen. Daher musste der Wert der monetären Beteiligungen aller AnteilseignerInnen geschätzt werden. Es zeigt sich, dass der Beteiligungswert trotz Schätzung stark und positiv mit dem individuellen Nettovermögen korreliert. *Über den generierten monetären Wert der Beteiligungen sollte also eine valide Stratifikation zwecks Oversamplings im Bereich hoher und höchster Vermögen in einer Hauptuntersuchung gelingen.*
4. Wir haben insgesamt 2.000 Adressen aus der Datenbank von AnteilseignerInnen an das Feldinstitut weitergegeben. Diese Adressen bildeten die Grundlage für 100 avisierte Interviews (Pretest). Davon konnten 532 Personen während der Feldphase tatsächlich kontaktiert werden, und 124 Personen wurden erfolgreich befragt. Dies entspricht einer Quote von ca. 23% der kontaktierten Personen. *Damit sollte unsere Personendatenbank von AnteilseignerInnen auch für eine Hauptuntersuchung mehr als ausreichend sein.*

5. Das durchschnittliche individuelle Nettovermögen der im Pretest befragten 124 Personen liegt bei ca. 10 Mio. Euro; im SOEP bei weniger als 100.000 Euro. Dies entspricht einem Verhältnis von 100 : 1. Das höchste individuelle Nettovermögen aus dem Pretest liegt bei über 200 Mio. Euro. Es ist also auch gelungen, mindestens eine Person aus dem Manager-Magazin zu befragen. *Der Pretest zeigt also, dass sich Personen mit hohen Vermögen nicht nur zielgenau ausmachen, sondern auch befragen lassen.*
6. *Insgesamt lassen die Ergebnisse aus dem Pretest erwarten, dass die Umsetzung des Konzepts im Rahmen einer Hauptuntersuchung die Forschungsdateninfrastruktur im Bereich Top-Vermögender nachhaltig verbessern würde.*
7. Weitergehende Informationen zur Feldphase können dem Methodenbericht des Feldinstituts Kantar Public entnommen werden, der diesem Bericht als Anhang beigelegt ist.

1 Einleitung

Die aktuelle Datenlage in Deutschland hat substanzielle Defizite hinsichtlich der Repräsentation von Personen bzw. Haushalten mit hohen Vermögen. Folglich ist die Aussagekraft empirischer Studien zur Vermögensverteilung eingeschränkt. Dies gilt umso mehr, als dass Dividenden und Zinsen seit der Einführung der Abgeltungsteuer im Jahr 2009 nicht mehr systematisch in der Einkommensteuerstatistik enthalten sind. Bis 2009 ließen sich die Anga-

ben zu den Kapitaleinkünften in der Lohn- und Einkommensteuerstatistik dazu nutzen, die Vermögenswerte am oberen Ende der Verteilung zumindest zu approximieren. Dementsprechend betont auch der fünfte Armuts- und Reichtumsbericht der Bundesregierung (5. ARB) die Notwendigkeit einer Verbesserung der Datengrundlage im Bereich hoher Vermögen (Bundesregierung, 2017).

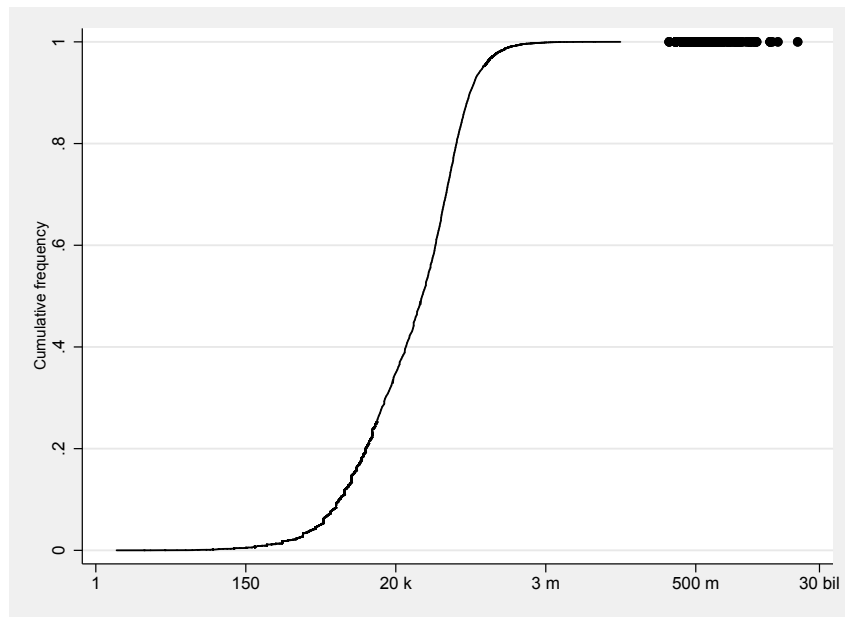
Tatsächlich finden sich in den drei großen repräsentativen Befragungen in Deutschland mit Informationen zu Privatvermögen² faktisch keine Haushalte mit einem Nettovermögen (Summe aus Geld- und Sachvermögen abzüglich Schulden) von 100 Mio. Euro oder mehr. Dabei leben laut Forbes-Liste 2017 rund 100 Milliardäre in Deutschland. Die Fallzahlen in den Befragungen reichen schon für Nettovermögen oberhalb von 10 Mio. Euro nicht für belastbare statistische Analysen aus (Westermeier and Grabka [2015]).

Damit gibt es in allen vorliegenden wissenschaftlichen Befragungen in Deutschland eine substanzielle Lücke in der Repräsentation im Top-Vermögensbereich. *Unter Top-Vermögende verstehen wir solche Personen, deren Vermögen mindestens im höchsten Bereich der Vermögensverteilung basierend auf existierenden Bevölkerungsbefragungen oder darüber liegt.*

Abbildung 1 zeigt die Datenlücke am Beispiel des SOEP. Dargestellt ist die kumulierte Dichte des individuellen SOEP-Nettovermögens laut; ferner die reichsten Personen laut Manager-Magazin. Die vertikalen Linien markieren die Datenlücke. Die Abbildung zeigt, dass die Fallzahl im SOEP im Bereich zweistelliger Millionenvermögen bereits sehr klein wird. Tatsächlich

²Hierbei handelt es sich um Private Haushalte und Finanzen (PHF), Sozio-oekonomisches Panel (SOEP) und Einkommens- und Verbrauchsstichprobe (EVS).

liegen die höchsten im SOEP erfassten Vermögen bei „nur“ rund 40 Mio. Euro. Laut Manager-Magazin gibt es aber alleine mindestens 100 Milliardäre in Deutschland. *Dies bedeutet, dass es eine eklatante Datenlücke im Bereich jenseits individueller Nettovermögen im niedrigen Millionenbereich gibt.*³



Anmerkung: Nettovermögen in Euro. Durchgezogene Linie: Individuelle SOEP-Nettovermögen (nur positive Nettovermögen); Punkte: Einzelpersonen aus Top-500 der Manager-Magazin Reichenliste 2014. Vermögen wurde auf log-Skala abgebildet. SOEP Daten für das Jahr 2012.

Abbildung 1: Kumulierte Dichte des Nettovermögens im SOEP vs. Manager-Magazin

Die schlechte Datenlage im Bereich hoher Vermögen hat mehrere Ursachen. Erstens ist die Wahrscheinlichkeit *per definitionem* gering, Top-Vermögende zufällig zu ziehen. Dementsprechend ist auch ihre Fallzahl in

³Diese Datenlücke wurde in der Vergangenheit durch Schätzungen unter Verwendung von Reichenlisten und der Pareto-Verteilungsannahme gefüllt. Danach gibt es einen log-log linearen Zusammenhang für Vermögenshöhe und empirischer Verteilung. Dieses Verfahren beruht auf verschiedenen Annahmen, die mit großen Unsicherheiten verbunden sind. Westermeier and Grabka [2015] zeigen, dass mit einer solchen Zuschätzung selbst einfache Trenderfassungen kaum möglich sind.

Zufallsbefragungen mit einigen tausend befragten Personen klein. Zweitens sinkt die Teilnahmebereitschaft systematisch mit der Höhe des Vermögens (Westermeier and Grabka [2015]). Dies reduziert den Anteil von Top-Vermögenden in Befragungen unter ihren tatsächlichen Bevölkerungsanteil. Drittens gibt es in Deutschland – im Unterschied zu vielen anderen Ländern – zu Privatvermögen keine Registerdaten, über die eine zielgenaue Identifikation Top-Vermögender für eine Befragung möglich wäre. Ein Oversampling (überproportionale Ziehung) von Personen mit Adressen in Regionen mit überdurchschnittlichem Steueraufkommen, wie etwa in der PHF-Befragung der Bundesbank, hat die Datenlage nicht substanziell verbessert. Eine nicht repräsentative Stichprobe anhand einfacher Verfügbarkeit (Convenience Sample), wie in der Studie Hochvermögende in Deutschland (HVID), erlaubt keine verallgemeinerbaren Aussagen.⁴

Ziel dieser Machbarkeitsstudie ist es daher zu prüfen, ob sich die Forschungsdateninfrastruktur im Bereich hoher Vermögen in Deutschland mit einem innovativen Ziehungs- und Befragungskonzept verbessern lässt. Dies gilt gerade auch in Hinblick auf zukünftige Armuts- und Reichtumsberichte.

Die grundlegende Idee unseres Konzepts fußt in der empirischen Beobachtung, dass fast alle vermögenden Personen zumindest einen Teil ihres Vermögens in Form von Beteiligungen an Unternehmen halten. Unternehmen wiederum veröffentlichen nicht nur Informationen über ihre Finanz- und Ertragslage, sondern auch über ihre Eigentümerstrukturen in Form von Namen, Adressen und Höhen der Beteiligungen. Unser Konzept nutzt diese

⁴Bei HVID wurde auch ein anderes Ziehungskriterium verwendet, nämlich ein Geldvermögen von mind. 1 Mio. Euro (Ströing et al. [2016]).

öffentlichen Informationen, um Personen in Deutschland mit Unternehmensbeteiligungen zu erfassen. Basierend auf dieser Grundgesamtheit wird eine nach dem Wert der Beteiligungen geschichtete Zufallsstichprobe zielgenau im Bereich hoher Vermögen gezogen und mit SOEP-Standard-Instrumenten befragt – insbesondere auch zu ihren Vermögen.

Mit der Implementierung des oben beschriebenen Konzepts ergeben sich verschiedene Herausforderungen und Arbeitsschritte:

1. **Erfassung von Personen mit substanziellen Unternehmensbeteiligungen.** Eine notwendige Bedingung für eine erfolgreiche Umsetzung des Konzepts war der Aufbau einer möglichst vollständigen Personendatenbank zu Namen, Adressen und Beteiligungen von Personen mit Wohnsitz in Deutschland, die nennenswerte Beteiligungen an Unternehmen weltweit halten. Dabei ging es nicht um Kleinaktionäre oder Eigentümer kleinerer Unternehmen (v.a. Personengesellschaften), sondern um Personen mit substanziellen finanziellen Beteiligungswerten. Hierzu haben wir die weltweite Unternehmensdatenbank *Orbis* des Dienstleisters Bureau van Dijk (BvD) verwendet, die Informationen zu mehr als 270 Millionen Unternehmen weltweit sowie deren Eigentümerstrukturen bereitstellt. Diese Eigentümerstrukturen können sehr komplex sein, weil viele Personen an einem Unternehmen beteiligt und Unternehmen zudem über Mutter-Tochter-Beziehungen verwoben sein können. Damit ergab sich technisch die Herausforderung, eine weltweite Unternehmensdatenbank mit komplexen Unternehmens-AnteilseignerInnen-Beziehungen in eine Personendatenbank zu über-

führen.

2. Identifikation von Personen mit hohem Vermögen. Die Grundidee des getesteten Konzepts besteht darin, dass Personen mit hohen monetären Unternehmensbeteiligungen auch vermögend sind. Allerdings beinhaltet Orbis keine direkten Informationen zu monetären Unternehmensbeteiligungen, sondern lediglich Unternehmensbeteiligungen in Prozent sowie Finanzkennzahlen der Unternehmen (Umsatz, Gewinn, etc.). Um die individuellen Beteiligungen zu monetarisieren, mussten wir daher zunächst den Unternehmenswert ökonometrisch schätzen. Damit haben wir anschließend jede prozentuale Beteiligung monetarisiert. So konnten wir dann die Summe aller Beteiligungen jeder Person bestimmen.

3. Kontaktierbarkeit und Befragungsbereitschaft von AnteilseignernInnen. Die AnteilseignerInnen aus der Personendatenbank müssen kontaktierbar,⁵ teilnahmebereit und gewillt sein, den SOEP-Fragebogen auszufüllen. Dazu gehören insbesondere Angaben zu Privatvermögen. Wir haben daher die Adressqualität geprüft und untersucht, wie zwei alternative Formen der Ansprache die Teilnahmebereitschaft beeinflussen.

Die zentralen Ergebnisse der vorliegenden Machbarkeitsstudie lassen sich wie folgt zusammenfassen:

⁵Für die Kontaktaufnahme ist die Aktualität und Qualität der Adressinformationen relevant. In der vorliegenden Studie bestand eine Herausforderung darin, dass nicht immer Privat- sondern teilweise nur Firmenadressen vorlagen, was eine persönliche direkte Kontaktaufnahme mit der Zielperson erschwert.

1. In der von uns erstellten Orbis-basierten Datenbank sind ca. 1,5 Mio. Personen mit deutschem Wohnsitz enthalten, die nennenswerte Unternehmensbeteiligungen im In- und/oder Ausland halten. *Diese Fallzahl erlaubt eine umfangreiche Implementierung des Konzepts im Rahmen einer Hauptuntersuchung.*
2. Die Vermögenden laut Manager-Magazin sind nahezu ausnahmslos in der Orbis-Datenbank wiederzufinden. *Dies lässt erwarten, dass sich mit unserem Konzept die Top-Vermögenden bis in den obersten Bereich der Verteilung erfassen lassen.*
3. Insgesamt haben wir 2.000 Namen und zugehörige Adressdaten aus unserer Personendatenbank an das Feldinstitut Kantar Public weitergegeben, das auch die SOEP-Erhebung durchführt. Dabei handelt es sich um eine stratifizierte Zufallsauswahl beschränkt auf fünf Regionen.⁶ Nachdem 124 Personen erfolgreich befragt wurden, wurde die Feldphase beendet, da die für den Pretest avisierte Stichprobengröße von 100 Personen bereits deutlich überschritten war. Die Teilnahmebereitschaft für die insgesamt 532 kontaktierbaren Fälle mit validen Adressinformationen lag damit bei 23% Prozent. Die Response-Rate in Verbindung mit der Größe der Personendatenbank (1,5 Mio. Personen) erlaubt damit auch eine groß angelegte Hauptuntersuchung. *Personen mit hohem Vermögen lassen sich also nicht nur zielgenau erfassen, sondern auch befragen.*
4. Das durchschnittliche Nettovermögen der im Pretest befragten 124

⁶Differenziert nach Stadt / Land und städtischer Randbereich.

Personen liegt bei über 10 Mio. Euro und übersteigt den SOEP-Durchschnitt um den Faktor 100. So liegt das höchste gemessene individuelle Netto-Vermögen im Pretest bei über 200 Mio. Euro. Damit ist es also gelungen, mindestens eine Person zu befragen, die Teil der Manager-Magazin-Liste sein könnte. *Die Ergebnisse aus dem Pretest suggerieren also, dass das Konzept die Lücke in wissenschaftlichen Befragungen im Bereich hoher Vermögen zu schließen vermag.*

5. Die geschätzten Beteiligungswerte der Befragten aus unserer Personendatenbank korrelieren stark positiv mit dem individuellen Nettovermögen, das sie im Interview angegeben haben. *Dies erlaubt in einer Hauptuntersuchung unter Verwendung der individuellen Beteiligungswerte ein effektives und damit vergleichsweise kostengünstiges Oversampling am oberen Ende der Vermögensverteilung.*

Insgesamt verspricht das Konzept, die systematische Untererfassung von Top-Vermögenden in Befragungen zu beheben – nicht nur in Deutschland. Eine größer angelegte Hauptuntersuchung dürfte, gerade als Panel angelegt, wertvolle Einsichten nicht nur zur Vermögenskonzentration in Deutschland liefern. Zu erwarten sind auch neue Erkenntnisse hinsichtlich folgender Themen: inter- bzw. intragenerationale Vermögensübertragungen; Vermögensaufbau im Lebenszyklus und seine Determinanten; Vermögenszusammensetzung; Persönlichkeitsmerkmale von Vermögenden; soziales Engagement, etc. Ferner dürfte eine Hauptuntersuchung auch die Dateninfrastruktur im Bereich hoher Einkommen verbessern, da Einkommen und Vermögen hoch positiv korreliert sind.

Die nachfolgenden Teile des Berichts gliedern sich wie folgt: Kapitel 2 erläutert die Identifikation der Grundgesamtheit. Dazu werden die Orbis-Unternehmensdatenbank, die Transformation der Daten in eine Personen-datenbank und die Bewertungsmethode der Unternehmensbeteiligungen beschrieben. In Kapitel 3 wird die Eignung der Orbis-Datenbank hinsichtlich der Repräsentation der Unternehmen und der Hochvermögenden sowie der verwendeten Methode zur Unternehmensbewertung diskutiert. Stichprobenziehung und Gewichtung werden in Kapitel 4 erläutert. Schließlich werden in Kapitel 5 die Ergebnisse der Pretest-Befragung vorgestellt und diskutiert. Kapitel 6 zieht ein Fazit aus der Machbarkeitsstudie.

2 Identifikation der Grundgesamtheit

Grundlage unseres Konzepts ist eine möglichst vollständige Personendatenbank zu Namen, Adressen und Beteiligungen von Personen mit Wohnsitz in Deutschland, die „nennenswerte“ Beteiligungen an Unternehmen weltweit besitzen. Diese Beteiligungen sollen hoch sein und ein hohes Privatvermögen erwarten lassen. Es geht also nicht um Eigentümer von (kleinen) Personengesellschaften etwa im Dienstleistungssektor oder Handwerk. Es geht auch nicht um Personen, die kleine Anteile an großen Unternehmen halten (Kleinaktionäre). Es geht um Personen mit substanziellen Beteiligungen an hohen Firmenwerten. Für diese Personengruppe haben wir unter Verwendung der Unternehmensdatenbank Orbis des Dienstleisters Bureau van Dijk (BvD) eine Personendatenbank aufgebaut.

Die nachfolgenden Abschnitte erläutern die Orbis-Datenbank, die Trans-

formation dieser in eine Personendatenbank von AnteilseignerInnen, sowie die monetäre Bewertung ihrer Beteiligungen an Unternehmen weltweit.

2.1 Die Orbis-Datenbank

Orbis stellt umfassende Informationen zu mehr als 270 Millionen Unternehmen einschließlich Banken und Versicherungen weltweit bereit. Diese basieren vor allem auf den veröffentlichten Bilanzen der Unternehmen. Verzeichnet sind diverse monetäre Finanzkennziffern, wie z.B. Umsatz, Eigenkapital oder Fremdkapital, ebenso wie Informationen zur Zahl der Mitarbeiter, zum Firmensitz oder den Beteiligungsstrukturen.⁷ Die Daten werden von Länderspezifischen Dienstleistern zugeliefert; für Deutschland von der Organisation Creditreform. Creditreform wiederum recherchiert im Unternehmensregister und nutzt darüber hinaus freiwillige Unternehmensauskünfte.⁸ Die in Orbis dokumentierten Unternehmensbilanzen- und -abschlüsse beinhalten auch Namen- und Adressdaten von Personen mit „nennenswerten“ Beteiligungen – solche im Ausmaß von mindestens 0,01 Prozent. Dieser Personenkreis ist die Zielpopulation unseres Konzepts.

Es sei angemerkt, dass die Freien Berufe und Kleingewerbetreibende nicht systematisch in Orbis erfasst sind, da diese weder zur doppelten Buchführung verpflichtet sind noch eine jährliche Bilanz erstellen müssen.⁹ Das Fehlen dieser Personengruppe ist für unser Vorhaben nicht von Bedeutung, da hierin kaum Top-Vermögende zu erwarten sind.

⁷ Vgl. Orbis Homepage.

⁸ Vgl. Handbuch zur deutschen BvD-Unternehmensdatenbank „Markus“.

⁹ Laut Statista gibt es ca. 1,4 Mio. Freiberufler in Deutschland.

2.2 Transformation der Unternehmens- in eine Personendatenbank

Orbis ist als Unternehmensdatenbank konzipiert. Zum Beispiel lässt sich aus Orbis direkt der Umsatz der Alpha-Omega AG¹⁰ oder ihre Mitarbeiterzahl abrufen ebenso wie die Struktur der Eigentümer.

Für unser Konzept wird dagegen eine Datenbank auf Personenebene, auf Ebene der AnteilseignerInnen, benötigt. Eine Aufgabe bestand somit darin, die Unternehmensdatenbank Orbis in eine Personendatenbank zu transformieren. Diese sollte alle AnteilseignerInnen mit Wohnsitz in Deutschland, ihren Namen und Adresse sowie ihre Beteiligungen an Unternehmen weltweit enthalten.

Schematisch erfolgte die Transformation der Datenbank in drei Schritten:

1. Orbis umfasst mehrere Datenbanken, von denen wir im ersten Schritt die so genannte Orbis-Kontaktdatenbank (`Contacts`) verwenden, um alle Personen mit Wohnsitz in Deutschland zu extrahieren, die aktuell Beteiligungen an mindestens einem aktiven Unternehmen weltweit halten. Hierzu filtern wir (`Search strategy`) über folgende Merkmale:

- `Contact gender: Male, Female, Unspecified`
- `Contact type: Individual`
- `Country of residence: Germany`
- `Current position: Shareholder not being a director in the same company, Shareholder also being a director in the same company`

¹⁰Fiktiver Firmenname.

- Legal status: Active companies (including those with unknown status)

Ergebnis dieser Abfrage ist allerdings eine Datenbank, die sowohl die Zielgruppe der AnteilseignerInnen umfasst, allerdings auch Personen ohne Beteiligungen, die lediglich eine offizielle Position in Unternehmen (z.B. GeschäftsführerIn) bekleiden.¹¹ Solche Personen sind nicht unsere Zielgruppe und wurden daher ausgeschlossen.

Die grundlegende Komplikation des ersten Arbeitsschritts besteht darin, dass die Bedienoberfläche von Orbis nur Abfragen mit maximal 100 Einträgen erlaubt. Die Bedienoberfläche hat auch keine Programmierschnittstelle, die ein simples Extrahieren der Objekte erlauben würde. Deshalb haben wir zur Automatisierung der Abfragen ein Script erstellt, das einen menschlichen Orbis-Nutzer imitiert, der per Maus- und Tastatureingabe die Adressen herunterlädt.¹²

2. Jedes Unternehmen in Orbis hat eine eindeutige Identifikationsnummer (BvD ID number). Anhand derer konnten Informationen wie Umsatz, Mitarbeiterzahl oder Beteiligung in Prozent an den in Schritt 1 gelisteten Unternehmen aus der Datenbank (Companies) extrahiert werden.¹³

¹¹Es gibt Personen mit offiziellen Positionen, die entgegen der Vorgaben in der Datenabfrage keine Beteiligungen besitzen.

¹²Die BvD-Bedienoberfläche ist eine Webanwendung in HTML und JavaScript. An dieser Stelle konnte daher auf das Python-Paket Selenium zurückgegriffen werden, das in der Softwareentwicklung zum automatisierten Testen von Webanwendungen genutzt wird. Auf diese Weise war es möglich, die in HTML-Tabellen dargestellten Adressen als HTML-Quelltext herunterzuladen. In einem nächsten Schritt extrahiert das Python-Skript die Adressen aus dem HTML-Quelltext und wandelt sie in ein übliches Tabellenformat (.csv) um.

¹³Die Unternehmens-IDs werden hierzu in die Suchmaske der Unternehmensdatenbank

3. Die Personendatenbank aus Schritt 1 wird mit den Unternehmensinformationen aus Schritt 2 verknüpft.

Im Ergebnis beinhaltet unsere Personendatenbank die Beteiligungen in Prozent und Adressdaten von rund 1,5 Mio. AnteilseignerInnen; außerdem diverse Finanzkennzahlen zu den jeweiligen Unternehmen.

2.3 Monetarisierung der prozentualen Beteiligungen

Für eine zielgenaue Befragung von Top-Vermögenden mussten alle prozentualen Beteiligungen in unserer Datenbank monetarisiert und anschließend auf Personenebene kumuliert werden. Das Ergebnis dieses Arbeitsschritts ist ein Betrag je AnteilseignerIn (*kumulierte Beteiligung*), der den Wert der Beteiligungen dieser Person weltweit beschreibt. Anschließend wurden die AnteilseignerInnen anhand ihrer Beträge absteigend geordnet, um später eine stratifizierte Zufallsstichprobe zu ziehen und zu befragen.

Bei der Monetarisierung waren nicht nur die direkten Beteiligungen zu bewerten, sondern auch etwaige Anteile an Tochterunternehmen. In der Orbis-Unternehmensdatenbank bedeutet dies, dass zwischen dem Wert der direkten Unternehmensbeteiligungen (`Shareholder -- Direct`) und ihrem Gesamtwert unter Einschluss indirekter Beteiligungen (`Shareholder -- Total`) zu unterscheiden war. Diese Differenzierung ist wichtig, weil Ab-

eingesehen. Hierbei gibt es technische Beschränkungen, weshalb die IDs in vier Blöcke zu je 500.000 Unternehmen aufgeteilt wurden. Jeder Block definiert eine einzelne Suchstrategie (`Search Strategy`), die dann in der Folge in einem booleschen Ausdruck mit „oder“ verknüpft werden können. Die relevanten Bilanz- und Erfolgskennziffern sowie AnteilseignerInnen-Informationen werden exportiert, wobei entsprechend Kalemli-Ozcan et al. [2015] bei Finanzvariablen ein `formatted export` durchgeführt wird. Der Output wird in mehreren Textdateien abgespeichert.

schlüsse teils konsolidiert, teils nicht konsolidiert vorliegen (`Consolidation code`). Konsolidiert bedeutet, dass in der Bilanz des Mutterunternehmens die Werte der Tochterunternehmen aufgeführt sind.

Ist der Abschluss konsolidiert, reichen der Prozentsatz der direkten Beteiligung sowie der Unternehmenswert aus der konsolidierten Bilanz aus, um den Wert der Beteiligung zu bestimmen. Ist der Abschluss nicht konsolidiert, ermittelt sich der Wert der Beteiligung aus der direkten Beteiligung am Mutterunternehmen (multipliziert mit dem nicht konsolidierten Wert desselben) *zuzüglich* der prozentualen Beteiligungswerte an den Töchtern.¹⁴

Abbildung 2 beschreibt schematisch ein Fallbeispiel, in dem Person 1 zehn Prozent der Anteile von Unternehmen 1 und 20 Prozent der Anteile von Unternehmen 2 hält. Ferner hält Unternehmen 2 eine 50-prozentige Beteiligung an Unternehmen 1. Mit W_1^b sei der Wert von Firma 1 und W_2^b der Wert von Firma 2 mit $b \in (\textit{konsolidiert}, \textit{unkonsolidiert})$ angegeben. Wenn die Abschlüsse der Firmen unkonsolidiert sind, ergibt sich der Gesamtwert der Beteiligungen, W , als $W = (0,1 + 0,5 \times 0,2) \times W_1^u + 0,2 \times W_2^u$. Sind die Abschlüsse konsolidiert, ist also der Wert der Beteiligung von Unternehmen 2 an Unternehmen 1 in die Unternehmensbewertung von Unternehmen 2 eingeflossen, so beträgt der Gesamtwert $W = 0,2 \times W_2^k$. Anderenfalls würde eine Doppelbewertung erfolgen.

Der Wert einer Beteiligung an einem börsennotierten Unternehmen entspricht der prozentualen Beteiligung multipliziert mit der Marktkapitali-

¹⁴Bei der Erstellung des Datensatzes für die Pretest-Stichprobenziehung haben wir aus Zeitgründen nur zwei Beteiligungsebenen berücksichtigt. Später haben wir dann fünf Beteiligungsebenen berücksichtigt.

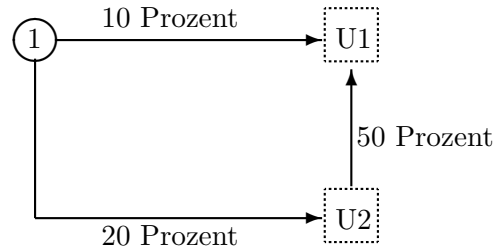


Abbildung 2: Schematische Darstellung: Beteiligungen bei konsolidierten und nicht konsolidierten Abschlüssen

sierung des Unternehmens. Die überwiegende Zahl der Unternehmen in Deutschland ist aber nicht börsennotiert. Daher benötigten wir ein einheitliches Bewertungsverfahren für alle Unternehmen. Dieses erfolgt über den Unternehmensumsatz.

Da Umsätze im Vergleich zu anderen Finanzkennzahlen für sehr viele Unternehmen vorliegen, ermöglicht dieses einfache Verfahren eine umfassende und einheitliche Form der Bewertung der Unternehmen und damit der *kumulierten Beteiligungen*.

Der Umsatz eines Unternehmens entspricht zwar nicht dem Unternehmenswert, ist mit diesem aber hoch positiv korreliert. Außerdem ist eine exakte Bewertung in Euro nicht notwendig, da die AnteilseignerInnen lediglich sortiert und ihr eigentliches Vermögen nachfolgend in der Befragung mit den SOEP-Instrumenten erhoben wurde.

Da Unternehmen zu unterschiedlichen Zeitpunkten ihre Bilanzkennziffern offenlegen, verwenden wir den aktuellsten Umsatz laut Orbis mit einem maximalen Verzug von fünf Jahren (also der aktuellste Umsatz der Jahre

2012 bis 2018). Liegt kein Umsatz vor,¹⁵ so wird dieser in einem zweistufigen Verfahren imputiert:

1. Für Unternehmen mit Angaben zum Umsatz, haben wir diesen ökonomisch unter Verwendung verschiedener Bilanzangaben erklärt. Dazu zählen der letzte verfügbare Umsatz im Zeitraum vor 2012, die letzte verfügbare Mitarbeiterzahl, die letzte verfügbare Bilanzsumme, sowie die Anzahl der Tochtergesellschaften und AnteilseignerInnen.
2. Für Unternehmen, deren Umsatz fehlt aber die erklärenden Variablen vorliegen, haben wir basierend auf dem obigen Modell den Umsatz vorhergesagt.
3. Fehlten sowohl Umsatz als auch erklärende Variablen, haben wir ein randomisiertes Verfahren implementiert, bei dem diesen Unternehmen eine gleichverteilte diskrete Zufallsvariable von 1 bis 100 zugeschrieben wurde. Anhand dieser wurde der Umsatz aus den 100 Perzentilen der beobachteten Verteilung der Umsätze bis 700.000 Euro zugewiesen. 700.000 Euro ist die Umsatzschwelle von Kleinstkapitalgesellschaften (§ 267a HGB), die ihnen erlaubt die ansonsten geltenden Offenlegungsvorschriften in geringerem Umfang nachkommen zu müssen.

Aus der Orbis-Datenbank haben wir mit der in Abschnitt 2.2 beschriebenen Vorgehensweise mehr als 1,5 Mio. AnteilseignerInnen extrahiert, die an mehr als 1,6 Mio. Unternehmen weltweit beteiligt sind. Von diesen Unternehmen haben rund 93 Prozent eine deutsche Identifikationsnummer. Dies

¹⁵Kleinstkapitalgesellschaften mit Sitz in Deutschland (§264 HGB) müssen laut § 326(2) HGB ihrer Offenlegungspflicht nur durch *Hinterlegung* beim Bundesanzeiger nachkommen.

ist der aus der Literatur bekannte „Home Country Bias“ (vgl. Coval and Moskowitz [1999]).

3 Eignung der Orbis-Daten für die Ziehung einer Hochvermögenden-Stichprobe

Die aus Orbis extrahierten Daten eignen sich für die zielgenaue Ziehung einer Hochvermögenden-Stichprobe, wenn folgende Aspekte zutreffen:

1. *Repräsentation von Unternehmen.* Sollen Top-Vermögende über ihre Beteiligungen an mittleren und großen Unternehmen erfasst werden, so ist es wichtig, dass Unternehmen dieser Größenordnung in der Orbis-Datenbank hinreichend repräsentiert sind. Um dies zu überprüfen, haben wir verschiedene Kennzahlen (z.B. Anzahl der Unternehmen nach Umsatz oder Rechtsform) der Orbis-Unternehmensdatenbank mit externen Datenquellen wie Unternehmensregister, Umsatzsteuer- oder Gewerbesteuerstatistik abgeglichen.
2. *Repräsentation von Top-Vermögenden.* Wenn es zutrifft, dass Top-Vermögende Unternehmensbeteiligungen halten, dann sollten sie auch als AnteilseignerInnen in Orbis zu finden sein. Um dies zu überprüfen, haben wir die 100 reichsten Menschen bzw. Familien laut Manager-Magazin in der von uns erstellten Personendatenbank recherchiert.
3. *Eignung des Konzepts zur Unternehmensbewertung.* Wie oben beschrieben, haben wir Unternehmenswerte und kumulierte Beteiligungen über die Umsätze der Unternehmen bestimmt. Dieses Bewertungs-

verfahren haben wir validiert. Zum einen haben wir alle AnteilseignerInnen anhand ihrer kumulierten Beteiligungen absteigend sortiert. Reiche Personen laut Manager-Magazin sollten sich weit oben in dieser Rangfolge positionieren. Zum anderen sollten die Privatvermögen derjenigen Personen, die mit SOEP-Instrumenten befragt wurden, hoch und positiv mit ihren kumulierten Beteiligungen korrelieren.

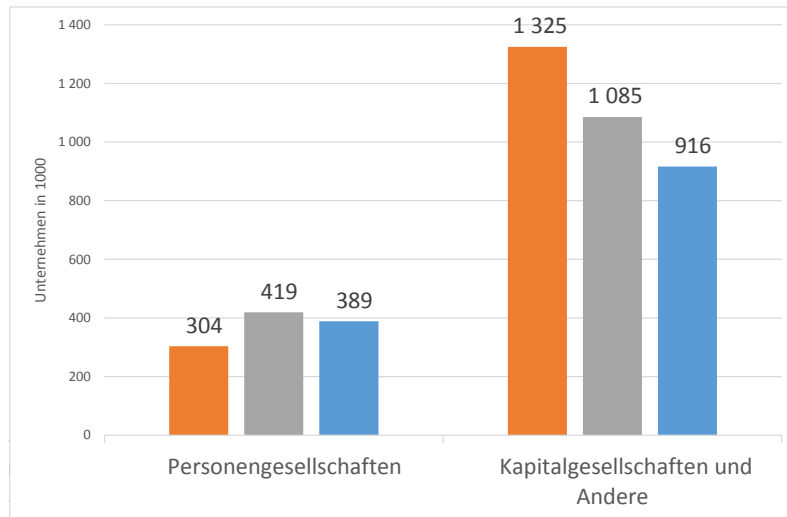
Nachfolgend zeigen wir, dass das Konzept eine zielgenaue repräsentative Stichprobenziehung von Top-Vermögenden erlaubt.

3.1 Repräsentation von Unternehmen in Orbis

Für die Einschätzung der Repräsentation von Unternehmen in Deutschland in der Orbis-Unternehmensdatenbank vergleichen wir in einem ersten Schritt die Fallzahlen in Orbis mit der Gewerbesteuerstatistik bzw. dem Unternehmensregister.

Die Fallzahlen zu Personen- und Kapitalgesellschaften laut Orbis entsprechen annähernd denen der Gewerbesteuerstatistik und des Unternehmensregisters. Abbildung 3 zeigt, dass Orbis rund 300.000 Personengesellschaften abbildet, das Unternehmensregister rund 390.000 und die Gewerbesteuerstatistik rund 420.000. Die geringe Fallzahl in der Orbis-Datenbank lässt sich insbesondere durch das Fehlen der „Gesellschaften bürgerlichen Rechts“ erklären, die nur im Gewerberegister geführt werden. Ferner ist zu beachten, dass wir in Orbis alle Unternehmen, die in mindestens einem der letzten fünf Jahre aktiv waren, berücksichtigt haben. Es kann daher sein, dass in unserer Datenbank Unternehmen enthalten sind, die heute nicht

mehr aktiv sind. Dies bedeutet aber nicht, dass die damaligen AnteilseignerInnen keine Privatvermögen akkumuliert haben.



Anmerkung: Eigene Berechnungen basierend auf Orbis (orange), Gewerbesteuerstatistik (grau), 2013 sowie Unternehmensregister (blau), 2015.

Abbildung 3: Zahl erfasster Unternehmen nach Rechtsform

Orbis erfasst rund 1,3 Mio. Kapitalgesellschaften, das Unternehmensregister rund 900.000 und die Gewerbesteuerstatistik rund 1,1 Mio. Die höhere Zahl von Kapitalgesellschaften laut Orbis im Vergleich zum Unternehmensregister erklärt sich dadurch, dass in diesem Register Unternehmen mit einem Umsatz von weniger als 17.500 Euro fehlen. Vermögensverwaltende GmbHs sind zwar Kapitalgesellschaften, können jedoch von der Gewerbesteuer ausgenommen sein, was eine weitere Erklärung für die unterschiedliche Fallzahl in Orbis und Gewerbesteuerstatistik ist.

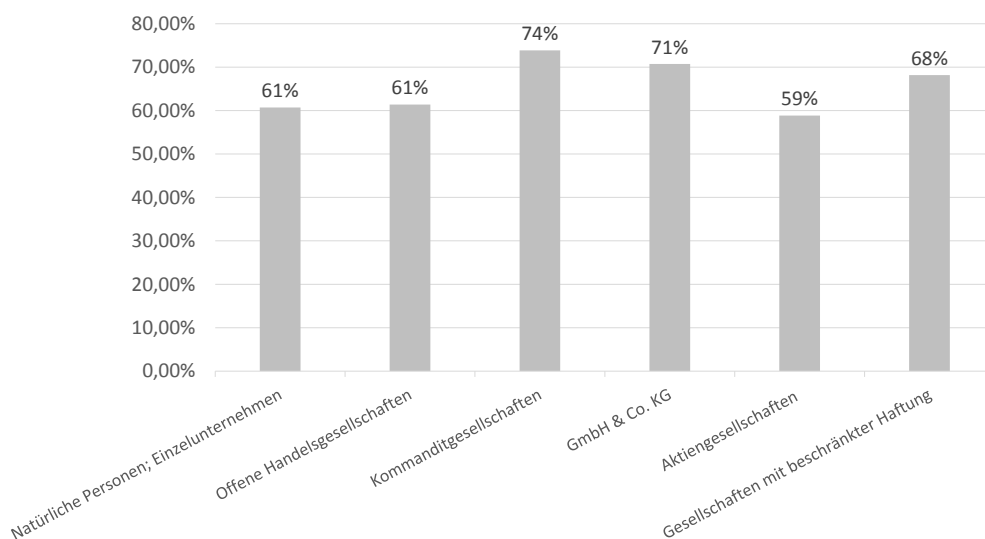
Info-Box: Offenlegungsvorschriften

Die im Handelsgesetzbuch (HGB) geregelten **Offenlegungsvorschriften** für Unternehmen hängen von der *Rechtsform* und der *Größe* des Unternehmens (gemessen an der Anzahl der Mitarbeiter, der Bilanzsumme oder dem Umsatz) ab. In Bezug auf die *Rechtsform* gilt, dass nur Kapitalgesellschaften und Personengesellschaften ohne persönlich haftenden Gesellschafter (z.B. GmbH und Co. KGs) ihren Jahresabschluss im Bundesanzeiger veröffentlichen müssen (§ 264 HGB). Sonstige Rechtsformen sind nur offenlegungspflichtig, wenn sie bestimmte Schwellenwerte für Umsatz, Mitarbeiterzahl oder Bilanzsumme überschreiten. Von der Offenlegungspflicht sind außerdem Tochtergesellschaften von konsolidiert abschließenden Unternehmen ausgenommen (§ 264(3) HGB). Vereinfachend gesprochen verringern sich die Erfordernisse zur Offenlegung mit der *Unternehmensgröße*. Nur große Kapitalgesellschaften müssen eine Gewinn- und Verlustrechnung und damit den Umsatz veröffentlichen (§ 275 HGB). Kleinere Kapitalgesellschaften (d.h. Jahresumsatz < 40 Mio. Euro) müssen ihren Umsatz nicht offenlegen (§ 326(1) HGB). Kleinstkapitalgesellschaften (Jahresumsatz < 700.000 Euro) sind vollständig von der Publizitätspflicht befreit, so dass hier lediglich eine Bilanz *hinter-* jedoch nicht *offengelegt* ist (§326(2) HGB).

Die Anzahl von kleinen Unternehmen in der Form von Einzelunternehmen bzw. natürlichen Personen ist in Orbis deutlich kleiner als in den beiden

anderen Quellen. Während Orbis nur ca. 140.000 Einzelunternehmen bzw. natürliche Personen erfasst, sind im Unternehmensregister und in der Gewerbesteuerstatistik ca. 2 Mio. gelistet. Diese Untererfassung ist für unser Vorhaben unproblematisch, da es unwahrscheinlich ist, dass Top-Vermögende *ausschließlich* Beteiligungen an Kleinbetrieben halten.

Aufgrund reduzierter Offenlegungspflichten für bestimmte Unternehmestypen kann es nicht überraschen, dass die Umsätze nur für rund 30 bis 40 Prozent der Unternehmen in Orbis direkt vorliegen. Abbildung 4 zeigt die Anteile der im Jahr 2017 aktiven Unternehmen ohne Umsatzangaben. Zudem besteht die Möglichkeit, dass Unternehmen ihre Bilanzen nicht rechtzeitig an den Bundesanzeiger übermitteln.

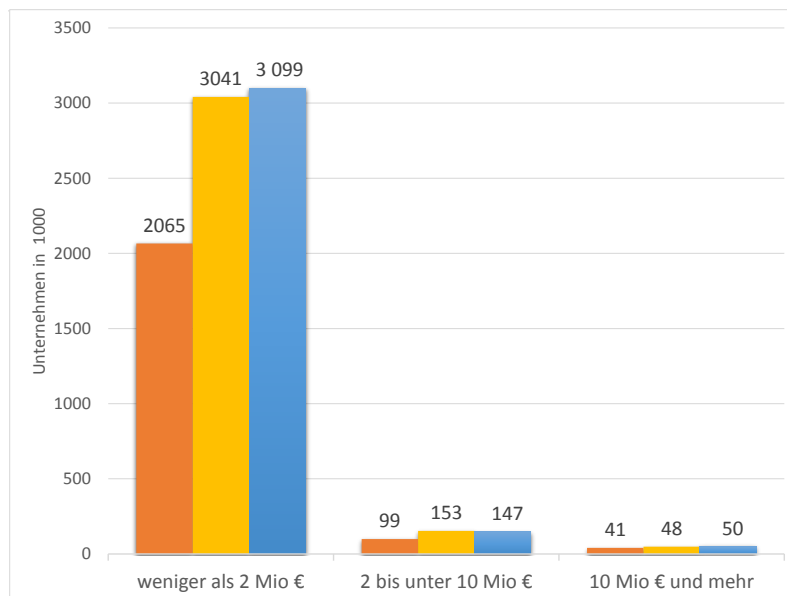


Anmerkung: Anteil der Unternehmen ohne Umsatzangabe der 2017 aktiven Unternehmen. Eigene Berechnungen basierend auf Orbis.

Abbildung 4: Anteile der Unternehmen ohne Umsatz in letzten fünf Jahren.

Fehlen Umsätze, haben wir diese für die nachfolgenden Auswertungen

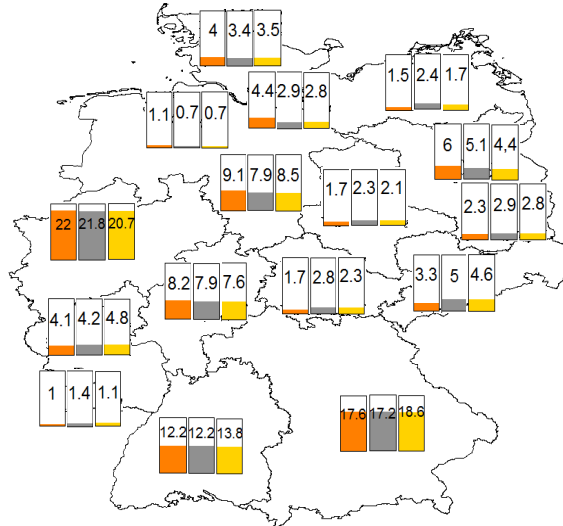
statistisch imputiert (vgl. Kapitel 2.3). Basierend auf dieser vervollständigten Datenbank konnte ein Abgleich mit externen Statistiken erfolgen. Abbildung 5 zeigt hohe Übereinstimmungen. So finden sich in Orbis und Unternehmensregister ähnliche Fallzahlen von Unternehmen mit einem Umsatz zwischen 2 und 10 Mio. wieder. Bei Umsätzen ab 10 Mio. Euro ist die Übereinstimmung noch höher. Die Kategorie unter 2 Mio. Euro Umsatz ist systematisch untererfasst. Erneut sei hier auf die geringere Fallzahl von Kleinbetrieben in der Orbis-Datenbank als Erklärung verwiesen.



Anmerkung: Eigene Berechnungen basierend auf Orbis (orange), Umsatzsteuerstatistik (Vorankmeldungen, gelb), 2013 sowie Unternehmensregister (blau), 2015.

Abbildung 5: Zahl erfasster Unternehmen nach Umsatz

Die regionale Verteilung der Unternehmen laut Orbis und Gewerbesteuerstatistik ist sehr ähnlich (Abbildung 6). Der Großteil der Unternehmen (ca. 22%) ist in Nordrhein-Westfalen ansässig, gefolgt von Bayern (ca. 17%) und Baden-Württemberg (ca. 12%).



Anmerkung: Eigene Berechnungen basierend auf Orbis (orange), Gewerbesteuerstatistik (grau), 2013 sowie Umsatzsteuerstatistik (Voranmeldungen, gelb), 2015.

Abbildung 6: Prozentualer Anteil erfasster Unternehmen je Bundesland

3.2 Repräsentation von Top-Vermögenden

Eine Übersicht zu den vermutlich 100-500 reichsten Deutschen wird jährlich vom Manager-Magazin publiziert. Bei den hierin angegebenen Vermögen handelt es sich um *Schätzungen* auf Basis frei zugänglicher Informationen. Hier verwenden wir die Daten aus der 2014er Ausgabe des Manager-Magazins.¹⁶

Trifft unsere Arbeitshypothese zu, wonach Vermögende auch Beteiligungen halten und ist zudem unsere Personendatenbank vollständig, sollten wir die reichsten Deutschen laut Manager-Magazin auch in unserer Personendatenbank finden.

Die Suche nach den reichsten Deutschen in unserer Personendatenbank

¹⁶Die Liste enthält 500 Personen/Familienverbände.

erwies sich als nicht trivial, da das Manager-Magazin teils Personen und teils Familienverbände listet. Bei einer Einzelperson befindet sich das geschätzte Vermögen auch im Besitz dieser Person. Bei Familien verteilt es sich auf die Familienmitglieder. So finden sich unter den 100 reichsten Deutschen laut Manager-Magazin 38 Einzelpersonen und 62 Familienverbände. Dabei wird Rang 1 von Stefan Quandt, Johanna Quandt und Susanne Klatten eingenommen, die unter anderem Beteiligungen an BMW und Altana halten, und deren *gemeinsames* Vermögen auf 31 Mrd. Euro geschätzt wird. Rang 5 wird von der Einzelperson Dieter Schwarz eingenommen, der Anteile z.B. an Lidl und Kaufland hält, und dessen *individuelles* Gesamtvermögen auf 14,5 Mrd. Euro geschätzt wird. Rang 10 wird von der „Familie Oetker“ eingenommen, deren Mitglieder aber nicht weiter spezifiziert werden.

Dies bedeutet, dass wir für jeden Familienverbund vor der Suche in der Personendatenbank die Familienbeziehungen recherchieren mussten; ferner bei häufigen Namen die Unternehmen, an denen die Personen beteiligt sein sollten. Aufgrund dieses hohen Aufwands haben wir diese Recherchen nur für die 100 reichsten Deutschen laut Manager-Magazin durchgeführt. Darunter fallen aufgrund teilweise identischer Vermögenswerte insgesamt 103 Personen und Familienverbände.

Zur Prüfung der Repräsentation der reichsten Deutschen in unserer Personendatenbank sind wir daher in zwei Schritten vorgegangen:

1. Über Wikipedia-Artikel sowie online zugängliche Zeitungs- und Zeitschriftenberichte, Gesellschafterverträge, etc. haben wir recherchiert, aus welchen Personen der Familienverband besteht. Die Identifikation

des „Familienoberhaupts“ war i.d.R. einfach und eindeutig. Bei (entfernten) Familienmitgliedern gibt es teilweise Unsicherheiten.

2. Anschließend haben wir alle Einzelpersonen und Familienmitglieder in der Personendatenbank recherchiert. Dazu haben wir Abgleiche über verschiedene Merkmale vorgenommen: Name, Vorname, Geburtsjahr und Unternehmensbeteiligung.

Insgesamt konnten wir den 100 reichsten Deutschen laut Manager-Magazin 404 Personen aus der Personendatenbank zuordnen. Bei 101 von 103 Personen bzw. Familienverbänden konnten wir mindestens ein Familienmitglied in Orbis identifizieren.¹⁷ Die verbleibenden zwei nicht identifizierten Fälle erklären sich durch einen Todesfall und einen Familienverbund (Familie Engelhorn), die im Manager-Magazin mehrere Ränge einnimmt, deren Mitglieder aber nicht eindeutig zuordenbar sind.¹⁸

Insgesamt zeigen diese Ergebnisse, dass wir mit unserer Personendatenbank die reichsten Deutschen laut Manager-Magazin quasi vollständig abdecken.

3.3 Validität der geschätzten kumulierten Beteiligungen

In den Abbildungen 7 und 8 sind die empirischen Verteilungen der kumulierten Beteiligungen aller 1,5 Millionen deutschen AnteilseignerInnen dargestellt. Rot eingefärbte Kreise zeigen, wo sich die 100 reichsten Deutschen

¹⁷Davon leben laut Orbis 88 Personen in Deutschland.

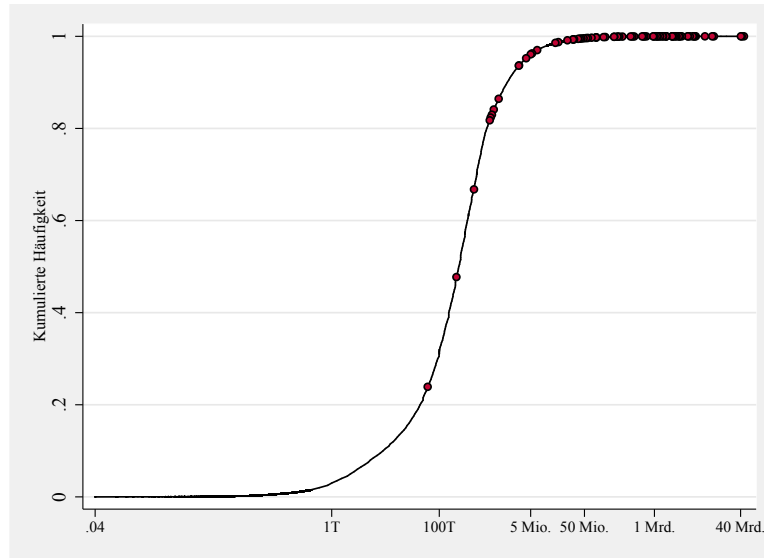
¹⁸Von den 404 Einzelpersonen konnten wir 331 in unserer Personendatenbank identifizieren. Die Differenz ergibt sich unter anderem daraus, dass Personen bereits verstorben sind, Anteile über die Zeit veräußert wurden, etc. 39 dieser 331 Personen leben im Ausland, 292 im Inland, 281 sind als AnteilseignerInnen mindestens eines Unternehmens aufgeführt.

laut Manager-Magazin in dieser Verteilung positionieren. Dazu haben wir bei Familienverbänden die kumulierten Beteiligungen jedes in der Personendatenbank identifizierten Mitglieds aufsummiert und *einer die Familie repräsentierenden Person* zugewiesen. Während Abbildung 8 nur Einzelpersonen in Form roter Kreise darstellt, zeigen diese in Abbildung 7 darüber hinaus auch die repräsentierenden Personen der Familienverbände. In beiden Abbildungen finden sich jeweils zwei Grafiken: Die obere Grafik berücksichtigt alle Fälle, bei denen eine Verknüpfung (sicher oder unsicher) der Personendatenbank mit den reichsten Deutschen laut Manager-Magazin möglich war. Die untere Grafik berücksichtigt ausschließlich sicher verknüpfbare Fälle.¹⁹

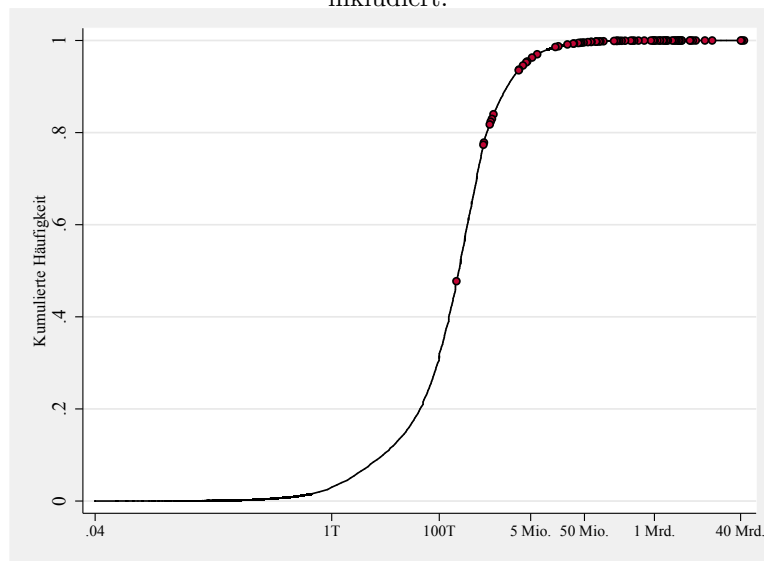
Es zeigt sich ein sehr einheitlicher und robuster Befund: Die reichen Deutschen laut Manager-Magazin konzentrieren sich in der Verteilung der kumulierten Beteiligungen im oberen Bereich. Dies gilt insbesondere bei der von uns präferierten ausschließlichen Betrachtung von Einzelpersonen. Denn bei diesen fallen im Vergleich zu Familienverbänden die Unsicherheiten (Aufteilung des Vermögens im Familienverbund; Abgrenzung des Familienverbunds; Präzision der Vermögensschätzung im Manager-Magazin) kleiner aus.

Tabelle 1 fasst diese grafischen Befunde anhand deskriptiver Maße zusammen. Dazu haben wir die kumulierten Beteiligungen aufsteigend sortiert und auf 100 Perzentile abgebildet. Jedes einzelne Perzentil umfasst genau

¹⁹Von den 103 reichsten Deutschen des Manager-Magazins konnten 81 in Orbis identifiziert werden. Davon waren fünf Zuordnungen unsicher (vgl. Spalte „Einzelpersonen & Familien*“). Betrachtet man lediglich die Einzelpersonen, so konnten 25 Einzelpersonen zugeordnet werden, wobei lediglich zwei Zuordnungen unsicher waren (vgl. Spalte „Einzelpersonen*“).



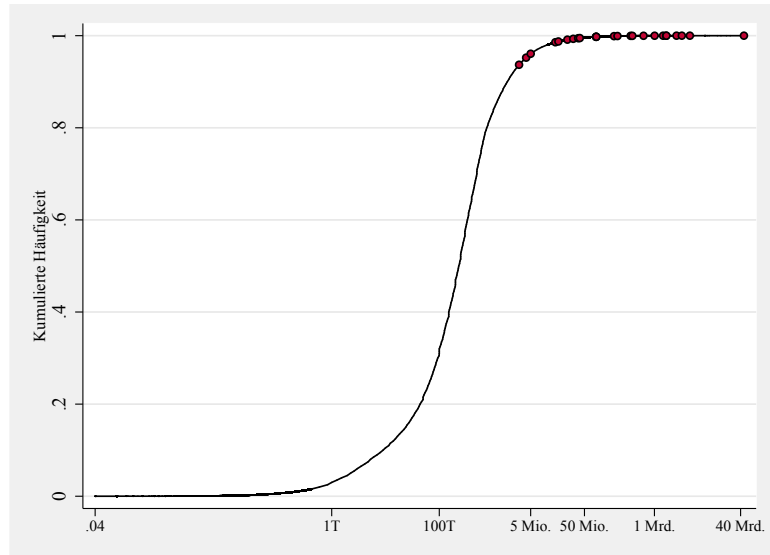
(a): Kumulierte Beteiligungen in Euro. Unsichere Matches inkludiert.



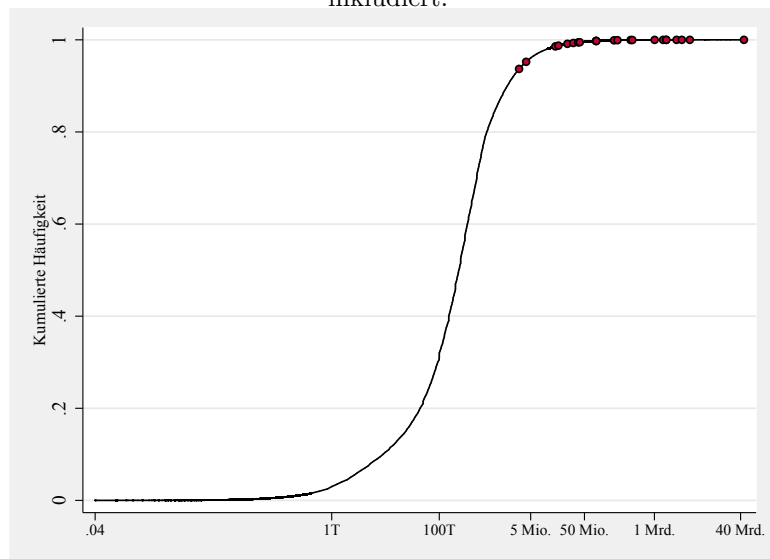
(b): Kumulierte Beteiligungen in Euro. Nur sichere Matches inkludiert.

Anmerkung: Datengrundlage sind unsere Personendatenbank und das Manager-Magazin (2014). Durchgezogene Linie: Empirische Dichte der kumulierten Beteiligungen; rote Punkte: gematchte Fälle aus Manager-Magazin; (a): unsichere Matches inkludiert; (b): nur sichere Matches.

Abbildung 7: Positionen der reichsten Deutschen laut Manager-Magazin in der Verteilung der kumulierten Beteiligungen – Einzelpersonen und Familien.



(a): Kumulierte Beteiligungen in Euro. Unsichere Matches inkludiert.



(b): Kumulierte Beteiligungen in Euro. Nur sichere Matches inkludiert.

Anmerkung: Datengrundlage sind unsere Personendatenbank und das Manager-Magazin (2014). Durchgezogene Linie: Empirische Dichte der kumulierten Beteiligungen; rote Punkte: gematchte Fälle aus Manager-Magazin; (a): unsichere Matches inkludiert; (b): nur sichere Matches.

Abbildung 8: Positionen der reichsten Deutschen laut Manager-Magazin in der Verteilung der kumulierten Beteiligungen – nur Einzelpersonen.

1% der AnteilseignerInnen. Im ersten Perzentil befinden sich die AnteilseignerInnen mit den geringsten und im 100sten Perzentil diejenigen mit den höchsten kumulierten Beteiligungen.

Die Zeile „Mittelwert“ gibt die durchschnittliche Perzentilposition der 100 reichsten Deutschen laut Manager-Magazin in der Verteilung der kumulierten Beteiligungen der Personendatenbank an. Je höher der Wert, desto stärker konzentrieren sie sich in den obersten Perzentilen. Tatsächlich ist der Mittelwert je nach betrachteter Gruppe zwischen 96,5 und 99,2. *Die 100 reichsten Deutschen positionieren sich also fast ausnahmslos in den höchsten Perzentilen der Verteilung der kumulierten Beteiligungen.*

Tabelle 1: Die 100 reichsten Deutschen in der Verteilung der kumulierten Beteiligungen

	Einzelperson*	Einzelperson	Einzelperson & Familien*	Einzelperson & Familien
Mittelwert	99,15	99,24	96,52	97,24
Std.abw.	1,65	1,58	11,27	7,77
Perzentil				
10	96,08	98,58	93,63	93,53
50	99,90	99,90	99,97	99,95
Fallzahl	25	23	81	76

Anmerkung: Eigene Berechnungen basierend auf Manager-Magazin und Orbis. * inklusive unsichere Matches. Datengrundlage sind Orbis (2014) und Manager-Magazin (2014).

4 Stichprobenziehung und Feldphase des Pretests

4.1 Stichprobenziehung

Die Stichprobenziehung für die Pretest-Befragung basiert auf der von uns erstellten Personendatenbank. Den Ziehungsrahmen (*Sampling Frame*) bilden

somit ca. 1,5 Mio. AnteilseignerInnen (siehe Kapitel 2) mit Unternehmensbeteiligungen weltweit und Wohnort in Deutschland. Vorab ausgeschlossen wurden Personen, die mit mindestens fünf anderen AnteilseignerInnen eine Adresse teilen. Denn dies ist ein klarer Hinweis, dass es sich nicht um eine Wohn- sondern eine Geschäftsadresse handelt²⁰. Insgesamt haben wir auf diese Art 1,4% der Fälle ausgeschlossen.

Die Zielpopulation der Stichprobe ergibt sich aus denjenigen Personen, die laut der Verteilung der kumulierten Beteiligungen (vgl. Kapitel 2.3) zum oberen Prozent der erwachsenen Bevölkerung Deutschlands gehören (ca. 660.000 Personen). Insofern wurden die knapp 1 Mio. Personen mit den niedrigsten kumulierten Beteiligungen ausgeschlossen.

Wie in bevölkerungsrepräsentativen Face-to-Face Studien üblich, wurde ein zweistufiger Stichprobenplan umgesetzt: In einem ersten Schritt wurde eine Zahl an Regionen (*Primary Sample Units, PSU*) und im zweiten Schritt Adressen innerhalb dieser Regionen (*Secondary Sample Units, SSU*) gezogen. Dieses zweistufige Verfahren erleichtert die Befragung aufgrund räumlich geklumpfter Adressen.

PSUs basieren auf Postleitzahlengebieten innerhalb von Bundesländern. Dabei wurden benachbarte Postleitzahlengebiete²¹ so zu PSUs zusammengefasst, dass in jeder PSU eine ähnlich hohe Zahl von Personen aus der Zielpopulation lebt. Infolgedessen variiert die räumliche Ausdehnung der PSUs

²⁰Zur näheren Prüfung, haben wir dem Erhebungsinstitut vor dem eigentlichen Pretest einige Adressen mit hohen Doppelungen übermittelt. Zwar handelte es sich bei einigen Adressen, insbesondere in Innenstadtlagen, tatsächlich um Wohnhäuser, in vielen Fällen handelte es sich aber um Geschäftsgebäude.

²¹Für ca. 0,8% der Personen aus der Zielpopulation ist in der Datenbank keine (korrekte) Postleitzahl hinterlegt. Sie fallen daher aus dem Ziehungsrahmen.

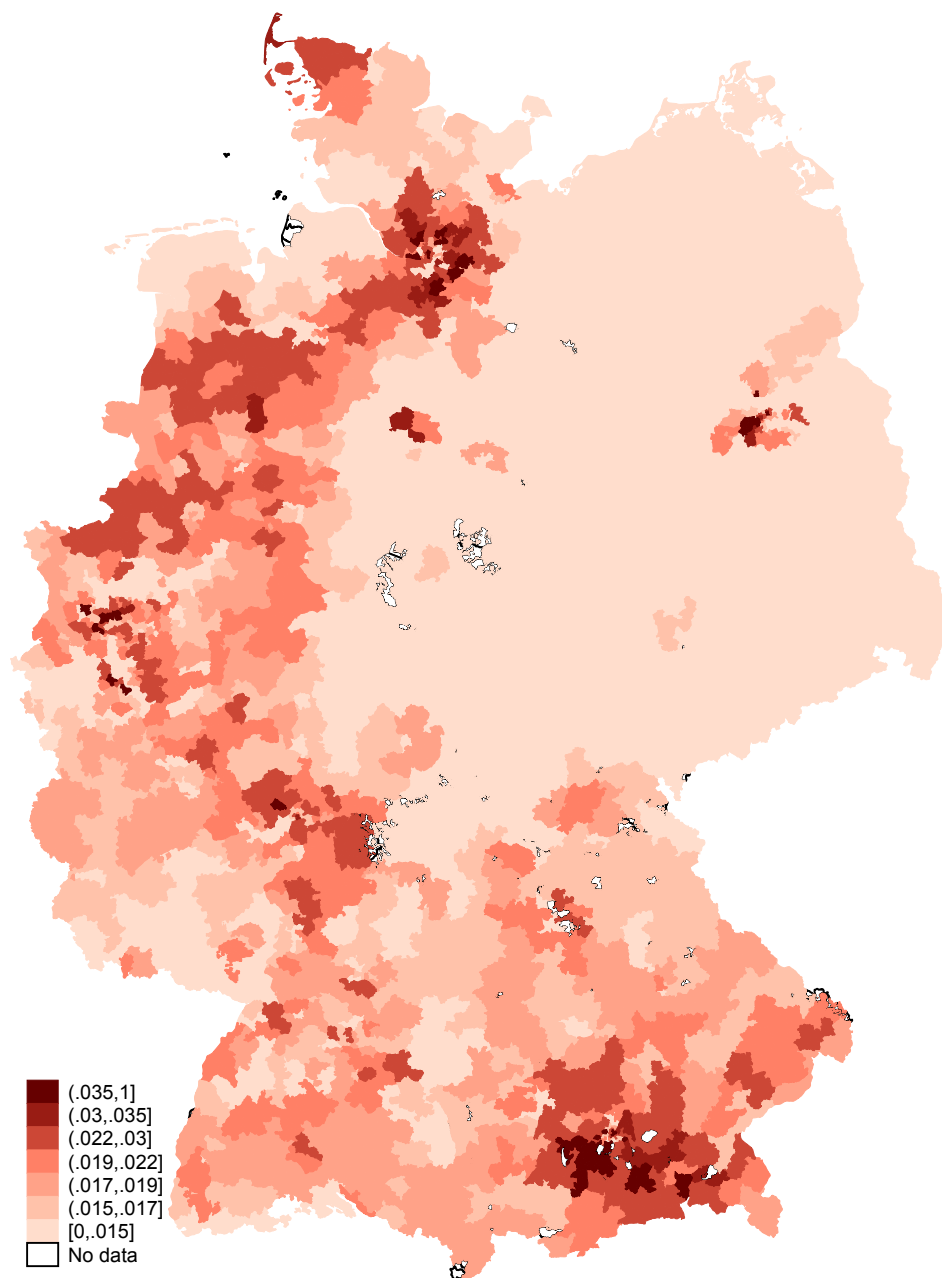
(siehe unten).

Insgesamt wurden 1.111 PSUs gebildet. Abbildung 9 veranschaulicht die regionalen Unterschiede bzgl. des Anteils der Zielpopulation an der Gesamtbevölkerung. Für den Pretest wurden jeweils 400 Adressen in fünf Regionen gezogen,²² die folgende Charakteristika aufweisen sollten:

1. *Verschiedenartige Gebietsstruktur.* Adressqualität, Antwortverhalten und sozioökonomische Zusammensetzung können sich regional und hinsichtlich des Umfelds unterscheiden. Daher wurden fünf möglichst heterogene Regionen ausgewählt. Unterschieden wird zwischen Innenstadt-, Stadtrand- und Vorortlage sowie ländlicher Raum. Ein Gebiet umfasst Innenstadt-, Stadtrand- und Vorortlagen sowie Stadtteile mit geringer Dichte der Zielpopulation.
2. *Geringe räumliche Ausdehnung.* Um die Wege für den jeweiligen Interviewer in einer Region zu minimieren, wurden für den Pretest Regionen mit besonders vielen Personen der Zielpopulation und geringer Ausdehnung ausgewählt.

Da entsprechend unserer bisherigen Erfahrungen davon auszugehen war, dass Kontaktierbarkeit und Teilnahmebereitschaft mit der Höhe der kumulierten Beteiligungen abnehmen, wurde eine innerhalb der Regionen nach der kumulierten Beteiligungshöhe geschichtete Zufallsziehung von Adressen durchgeführt. Dadurch wurde die Zielpopulation in drei Strata unterteilt:

²²Um zu gewährleisten, dass auch bei einer möglichen Haupterhebung noch hinreichend unbefragte Personen vorhanden sind, werden für den Pretest innerhalb dieser Regionen jeweils vier benachbarte PSUs zusammengefasst, aus denen die Befragungsadressen schließlich gezogen werden.



Anmerkung: Aus Darstellungsgründen wird auf eine Visualisierung der PSU-Grenzen verzichtet. Nebeneinanderliegende PSUs gleicher Färbung können daher nicht unterschieden werden. Ein Wert von 0.03 bedeutet 3%.

Abbildung 9: Anteil von Zielpersonen an der erwachsenen Gesamtbevölkerung auf Ebene der PSUs.

1/7 der gezogenen Adressen stammte aus dem unteren, 2/7 aus dem mittleren und 4/7 aus dem oberen Drittel der Verteilung der kumulierten Beteiligungen. Wir haben auch aus dem unteren Stratum gezogen, um die Zielgenauigkeit des Stratums zu prüfen. Zwischen den ausgewählten Gebieten unterscheidet sich die Verteilung auf die Strata, so dass die Ziehungswahrscheinlichkeiten der Bruttostichprobe über die Gebiete und Strata hinweg zwischen 5,3% und 34,3% variiert.

Um Erkenntnisse über Vor- und Nachteile einer auf die besondere Zielpopulation ausgerichtete Ansprache der Befragten zu gewinnen, wurde ein Experiment implementiert. Während die eine Hälfte der zu Befragenden das im SOEP übliche Informationsmaterial bekam, wurde die andere Hälfte explizit in ihrer Eigenschaft als Unternehmer angesprochen und zur Teilnahme motiviert. Bezüglich der Teilnahmebereitschaft zeigen sich nur geringe Unterschiede zwischen den beiden Varianten. Die höhere Teilnahmequote wurde allerdings mit der Standardansprache realisiert.

Von den 2.000 Adressen konnten etwa 13% nicht ermittelt werden, was eine gute Adressqualität für eine Stichprobe ist, die nicht auf Melderegisterdaten basiert (vgl. Kroh et al. [2015]). Teilweise hat die Adressaktualität Recherchen beim jeweiligen EMA notwendig gemacht hat. Diese waren auch zumeist erfolgreich. Ein ungewöhnlich hoher Anteil von 56% der 2.000 Fälle war während der gesamten Feldzeit nicht erreichbar - üblich sind hier Werte um 10%. Werden qualitätsneutrale Ausfälle berücksichtigt, wie etwa verstorbene oder ins Ausland verzogene Personen, verbleiben 532 Adressen, bei denen die Zielpersonen erfolgreich kontaktiert werden konnten.

Insgesamt wurden 124 erfolgreiche Interviews realisiert. Dies ergibt eine

unbereinigte Ausschöpfungsquote von 6%. Allerdings wurde die Feldphase vorzeitig beendet, da die vereinbarte Zahl erfolgreicher Interviews (100) überschritten war. Damit ist die angegebene unbereinigte Ausschöpfungsquote als Untergrenze für eine maximal erreichbare Quote anzusehen. Bezieht man die 124 Teilnahmen auf diejenigen 532 Personen, die während der Feldphase tatsächlich kontaktiert werden konnten, ergibt sich eine Quote von ca. 23%.

Erste Analysen zeigen, dass der Anteil der realisierten Interviews mit zunehmendem Wert der kumulierten Beteiligung (entsprechend den Strata) sinkt, mit dem Alter steigt und bei Männern höher ist als bei Frauen. Zwischen den einzelnen Regionen (und InterviewerInnen) gibt es nur geringe Unterschiede.

Damit sind Adressqualität wie Teilnahmebereitschaft ausreichend für eine größer angelegte Hauptuntersuchung.

4.2 Feldphase

Die Feldphase des Pretests begann am 06.11.2017 und endete am 02.02.2018. Die Interviews erfolgten computergestützt und persönlich (CAPI) durch das Feldinstitut Kantar Public. Die durchschnittliche Interviewdauer betrug 58 Minuten. Durch den Einsatz von Laptops bestand für die Befragten die Möglichkeit, gerade sensible Fragen zum Einkommen und Vermögen ohne direkte Einsicht des Interviewers zu beantworten. Der persönliche mündliche Kontakt war aufgrund der Befragungsdauer wichtig, um Abbrüche zu vermeiden.

5 Ergebnisse der Pretest-Befragung

Die nachfolgende Analyse hat zwei Ziele. Zum einen den Vergleich der Vermögenswerte aus Pretest und SOEP; zum anderen einen Vergleich der Vermögen über die drei Strata des Pretests. Dazu haben wir Pretest- und SOEP-Daten zusammengespield und vergleichend ausgewertet. Auf darüber hinausgehende inhaltliche Analysen, etwa zur Vermögensgenese oder der Portfoliozusammensetzung, haben wir verzichtet, da die Fallzahl aus dem Pretest keine generalisierbaren Aussagen erlaubt.

5.1 Vermögensmodul

Der im Pretest verwendete Fragebogen basierte auf dem regulären Personenfragebogen, der 2017 im SOEP eingesetzt wurde, einschließlich dem SOEP-Fragebogenmodul „Ihre persönliche Vermögensbilanz“. Damit war es möglich, Angaben zum Vermögen und anderen Charakteristika der Befragten aus SOEP und Pretest direkt zu vergleichen.²³ Für die vergleichenden Auswertungen von Pretest und SOEP wurden die zuletzt verfügbaren Angaben zur persönlichen Vermögensbilanz des Jahres 2012 verwendet.²⁴

²³Weiterhin wurden einzelne Fragen aus dem SOEP-Haushaltsfragebogen und des Fragebogens der HVID-Studie verwendet.

²⁴Das Modul „Ihre persönliche Vermögensbilanz“ wird im SOEP seit 2002 im Fünfjahresturnus erhoben. Darüber hinaus wurden im Jahr 2012 die Kategorien Fahrzeuge und Ausbildungskredite nicht erfragt. Diese sind allerdings quantitativ vernachlässigbar.

Das SOEP-Vermögensmodul umfasst insgesamt zwölf Vermögens- und Schuldenpositionen:²⁵

1. Wert selbst genutzter Immobilien
2. Restschuld auf Darlehen für selbst genutzte Immobilien
3. Wert sonstige Immobilien
4. Restschuld auf Darlehen auf sonstige Immobilien
5. Wert von Bausparverträgen
6. Höhe des Geldvermögens
7. Wert von Lebens- und privaten Rentenversicherungen
8. Wert eines Betriebs bzw. von Beteiligungen an Unternehmen
9. Höhe des Sachvermögens
10. Restschuld auf Konsumentenkredite
11. Wert von Fahrzeugen
12. Restschuld auf Ausbildungskredite

5.2 Statistische Imputation der Pretestdaten

Alle freiwilligen Bevölkerungsbefragungen sind mit Antwortverweigerungen konfrontiert. Diese Verweigerung kann vollständig sein (Unit non-response) oder einzelne Fragen betreffen (Item non-response). Item non-response findet man gerade bei sensiblen Fragen. Hierzu zählen auch Fragen zum Einkommen und Vermögen (vgl. R. Frick et al. [2010]). Auf eine Korrektur von Unit non-response mittels Ungewichtung wurde verzichtet, da die Ergebnisse des Pretests nicht generalisierbar sind. Da Item non-response aber zu

²⁵Dabei wird jeweils vorab eine Filterfrage gestellt, ob die Person Vermögen oder Schulden laut der angegebenen Position besitzt. Trifft dies zu, wird der Wert abgefragt.

systematischen Verzerrungen führen kann – gerade auch bei Angaben zum Einkommen und Vermögen – haben wir fehlende Werte statistisch imputiert.

Wie bei den Vermögen im SOEP verwenden wir zur Imputation das Verfahren der „Multiple Imputation by Chained Equations“ (MICE) (vgl. Royston and White [2011]). Damit können Kovariate verschiedener Skalenniveaus berücksichtigt und Unsicherheiten hinsichtlich der imputierten Werte abgebildet werden. Konkret basiert MICE auf einer Reihe von Regressionsmodellen, die simultan jede Variable mit fehlenden Antwortangaben mit anderen Variablen bedingen. Damit bleibt die Variabilität in der Datenstruktur (Vermögen) erhalten.

In der Pretest-Implementation von MICE beträgt der erklärte Anteil der Varianz an der Gesamtvarianz (R^2) je nach Vermögensaggregat bei den Filterfragen zwischen 12% und 57% und bei den Vermögenswerten zwischen 9% und 90%. Damit ist der Erklärungsgehalt des Modells insbesondere für einen Querschnittsdatensatz mehr als zufriedenstellend.

5.3 Vermögensvergleich zwischen SOEP und Pretest

Für den Vermögensvergleich zwischen SOEP und Pretest sowie die Vergleiche zwischen den drei Strata aus dem Pretest wurde das individuelle Nettovermögen verwendet.²⁶

Einen deskriptiven Vergleich der Vermögen in SOEP und Pretest erlaubt Tabelle 2. Die Spalten 2 und 3 beschreiben die Nettovermögensverteilungen unter Verwendung von Mittelwerten, Minimum, Maximum und von Perzen-

²⁶Die Ergebnisse für andere Vermögensaggregate sind qualitativ vergleichbar.

Tabelle 2: Individuelle Nettovermögen in Mio. Euro – SOEP vs Pretest

	Pretest				
	SOEP	Strata 1-3	Stratum 1	Stratum 2	Stratum 3
Minimum	-4	-2,5	-2,5	-1,2	-2,5
Mittelwert	0,09	10,3	4,2	3,3	17,4
Maximum	39,3	207	19,4	28,9	207
Perzentil					
25	0	0,7	0,5	0,5	1,5
50	0,02	2,3	2	1,1	3,6
75	0,1	6,6	6	3,3	16,2
90	0,2	23,4	10,8	7,6	52,7
95	0,3	42,6	16,8	11,8	89
99	0,8	156,6	19,4	28,9	207
Beobachtungen	25.803	124	23	41	60

Anmerkung: Eigene Berechnungen basierend auf SOEP 2012 (gewichtet) und Pretest (ungewichtet).

tilen. Der Vergleich zeigt, dass einzig das Minimum in SOEP und Pretest von gleicher Größenordnung ist, bei rund -4 Mio. Euro im SOEP und -2,5 Mio. Euro im Pretest.²⁷ Das durchschnittliche Nettovermögen im Pretest liegt bei über 10 Mio. Euro und damit mehr als hundert Mal höher als im SOEP (ca. 90.000 Euro). Tatsächlich liegen die Nettovermögen aus dem Pretest über die gesamte Verteilung hinweg weit über den SOEP-Werten. So liegt der Median des Pretests bei rund 2,3 Mio. Euro, während er im SOEP bei weniger als 20.000 Euro liegt. Der oberste Perzentilswert des Pretests liegt bei rund 157 Mio. Euro, im SOEP bei rund 838.000 Euro.

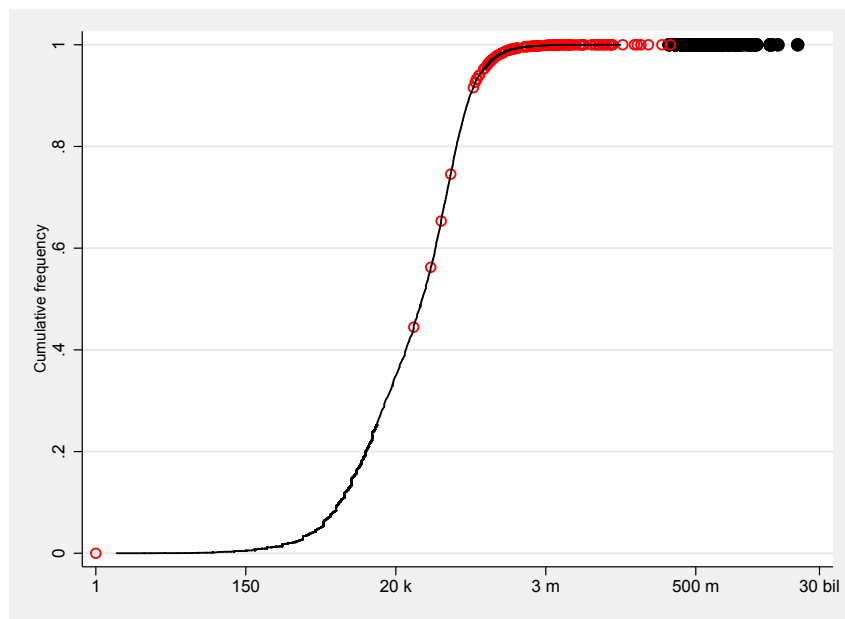
Es gelingt also mit dem Konzept, Personen mit hohen Vermögen zielgenau zu befragen. Tatsächlich liegt das Vermögen im 25. Pretest-Perzentil bereits auf dem Niveau des 99. SOEP-Perzentils.

Die Spalten 4 bis 6 der Tabelle 2 geben Aufschluss darüber, inwiefern die

²⁷Dass auch im Pretest insgesamt fünf Befragte ein negatives Nettovermögen haben, mag auf den ersten Blick überraschen. Allerdings ist auch bei AnteilseignernInnen nicht ausgeschlossen, dass sie sich verschulden, etwa um unternehmerische Aktivitäten zu finanzieren.

Schichtung über den Umsatz-Vermögensproxy im Pretest Personen in den oberen Bereichen der Vermögensverteilung erfolgreich differenziert. Insgesamt haben wir drei Schichten gebildet. Während sich die Vermögen in den Schichten 1 und 2 nicht systematisch unterscheiden, finden sich wesentlich höhere Vermögen in Schicht 3: Über die Perzentile hinweg liegen die Nettovermögen in Schicht 3 deutlich (Faktor 3 bis 10) über denen der beiden unteren Schichten. Der Mittelwert der Schicht 3 ist 200 Mal höher als im SOEP.

Die Stratifizierung scheint damit eine strukturierte Stichprobenziehung entlang der Datenlücke bestehender Befragungen zu ermöglichen.



Anmerkung: Durchgezogene Linie: Individuelle SOEP-Nettovermögen (nur positive Nettovermögen) mit Daten aus dem Jahr 2012; schwarze Punkte: Einzelpersonen aus Top-500 der Manager-Magazin Reichenliste 2014. Vermögen wurde auf log-Skala geplottet; rote Kreise: Beobachtungen aus Pretest. Vermögen wurde auf log-Skala geplottet.

Abbildung 10: Kumulierte Dichte des Nettovermögens im SOEP, Manager-Magazin und Pretest

Abbildung 10 zeigt noch einmal grafisch, wie erfolgreich die Pretest-Daten die Datenlücke schließen. Sie ergänzt Abbildung 1 um die Nettovermögen aus dem Pretest, die mit roten Kreisen dargestellt sind.

Während weniger als 1% der SOEP-Befragten zur Gruppe der Millionäre zählen, sind es rund 70% der Pretest-Befragten. 24 Pretest-Befragte verfügen über ein individuelles Nettovermögen von mindestens 10 Mio. Euro, sechs von über 50 Mio. Euro. Tatsächlich wurde auch mindestens eine Person mit rund 207 Mio. Euro befragt, ein Wert im Bereich der 500 reichsten Deutschen laut Manager-Magazin.

Damit verteilen sich die Pretest-Daten über die gesamte Datenlücke, was hoffen lässt, dass eine breit angelegte Hauptuntersuchung unter Verwendung des Konzepts diese Lücke tatsächlich schließen kann.

Die deskriptive Analyse prüft weder, ob die Unterschiede zwischen SOEP und Pretest (sowie zwischen den Pretest-Schichten) signifikant sind, noch kontrolliert sie für Kovariate. So könnte es sein, dass die Befragten im Pretest deshalb über höhere Vermögen verfügen, weil sie im Durchschnitt älter und höher gebildet sind als der SOEP-Durchschnitt oder häufiger männlichen Geschlechts sind.

Um die Differenz zwischen Pretest und SOEP zu schätzen, verwenden wir das folgende Modell:

$$NW_i = \alpha + \beta \times \mathbf{X}_i + \beta^{Pre} \times D_i^{Pretest} + \epsilon_i \quad (1)$$

mit der abhängigen Variablen, NW_i , dem logarithmierten Nettovermögen der Person i . Als Kovariate berücksichtigen wir:

1. Einen Vektor soziodemografischer Merkmale jeder befragten Person i , $\mathbf{X}_i$. Hierin enthalten sind Alter, berufliche Stellung und Geschlecht.
2. Eine 0-1-kodierte Dummy-Variable, $D_i^{Pretest}$, die angibt, ob die Person am Pretest oder am SOEP teilgenommen hat. Hat sie am Pretest teilgenommen, so nimmt der Dummy den Wert 1 an. Je größer der zu $D_i^{Pretest}$ gehörige Regressionskoeffizient, β^{Pre} , desto größer ist der Unterschied im durchschnittlichen bedingten Vermögen zwischen Pretest- und SOEP-Befragten und desto erfolgreicher ist die Stichprobenziehung im Sinne einer zielgenauen Befragung von Top-Vermögenden.

Eine weitere Spezifikation erlaubt es die Trennschärfe der Strata zu bestimmen. Sei K_i^{Strat} ein System kategorialer Dummy-Variablen mit $Strat = (1, 2, 3)$, wobei SOEP-Fälle mit 0 kodiert wurden. Die Schätzgleichung lautet:

$$NW_i = \alpha + \beta \times \mathbf{X}_i + \beta^{Strat} \times \mathbf{K}_i^{strat} + \epsilon_i \quad (2)$$

Je stärker die Koeffizienten β^{Strat} mit dem Stratum ansteigen, desto trennschärfer diskriminieren die Strata zwischen verschiedenen Vermögensbereichen.

Tabelle 3 fasst die Regressionsergebnisse zusammen. Dabei haben wir nur streng positive Vermögensangaben verwendet. Die Spalten *I* und *III* zeigen die Ergebnisse für die beiden Modellspezifikationen in der Basisvariante ohne soziodemografische Kontrollvariablen. Spalte *I* berücksichtigt nur

Tabelle 3: Regressionsergebnisse zur Höhe des Nettovermögens

	I	II	III	IV
Pretest	4,121*** (0,194)		1,652*** (0,187)	
<i>Stratum</i>				
1		4,037*** (0,339)		1,869*** (0,332)
2		3,291*** (0,401)		1,322*** (0,379)
3		4,723*** (0,224)		1,828*** (0,224)
Alter			0,173*** (0,004)	0,173*** (0,004)
(Alter) ²			-0,001*** (0,000)	-0,001*** (0,000)
Weiblich			-0,209*** (0,021)	-0,209*** (0,021)
<i>Beruf. Stellung</i>				
Angestellte			0,033 (0,030)	0,033 (0,030)
Selbstst. ohne Mitarbeiter			0,484*** (0,057)	0,486*** (0,057)
Selbst. mit 1-9 Mitarbeitern			1,096*** (0,065)	1,096*** (0,064)
Selbstst. mit mehr als 10 Mitarbeitern			2,003*** (0,127)	1,966*** (0,125)
Log Haushalts- nettoeinkommen			1,029*** (0,022)	1,029*** (0,022)
Konstante	10,58*** (0,013)	10,58*** (0,013)	-2,569*** (0,192)	-2,568*** (0,192)
<i>N</i>	19102	19102	19101	19101
Adj. <i>R</i> ²	0,032	0,033	0,374	0,374

Anmerkungen: Daten aus SOEP v.33 sowie Pretest; eigene Berechnungen.

Robuste Standardfehler in Klammern.

Abhängige Variable: Log des individuellen Nettovermögens mit imputierten Werten.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

den Dummy für den Pretest. Hier ist die Konstante positiv und signifikant. Ihr Wert entspricht $\exp(10,58) = 39.340$ Euro und beschreibt das durchschnittliche individuelle Nettovermögen im SOEP. Der Koeffizient für den Pretest-Dummy beträgt $\beta^{Pre} = 4,121$, was einem Durchschnittsvermögen im Pretest von $\exp(10,58 + 4,121) = 2.384.830$ Euro entspricht. Spalte *II* verwendet Dummies für die drei Strata des Pretests. Alle drei Strata-Koeffizienten sind signifikant und positiv. D.h., das durchschnittliche Nettovermögen ist in jeder Pretest-Schicht höher als der SOEP-Durchschnitt. Ferner liegt der Wert des Koeffizienten für das dritte Stratum deutlich über den beiden anderen Strata. Dies stützt den deskriptiven Befund, dass wir im Stratum 3 auch tatsächlich die höchsten Vermögen finden.

In den Spalten *III* und *IV* werden die erklärenden Variablen der Basisvariante um soziodemografische Variablen ergänzt. Die Ergebnisse zeigen, dass die Nettovermögen im Pretest auch nach Kontrolle beobachtbarer Charakteristika signifikant höher sind als im SOEP.²⁸ Ein einfaches Screening über die hier berücksichtigten Kontrollvariablen hätte folglich keinesfalls die Zielgenauigkeit des hier vorgeschlagenen Konzepts erreicht.

Die Regressionsergebnisse zeigen damit, dass die Vermögen im Pretest nicht nur signifikant und quantitativ weit höher sind als im SOEP, sondern auch, dass die Stratifikation eine gezielte Steuerung der Stichprobenziehung entlang des gesamten Vermögensspektrums der Datenlücke erlaubt.

²⁸Die Schätzergebnisse bzgl. der soziodemografischen Charakteristika entsprechen den bestehenden Befunden aus der Literatur: Das Nettovermögen steigt mit dem Alter und dem Haushaltsnettoeinkommen. Männer verfügen über höhere Vermögen als Frauen.

6 Fazit der Studie

Ein breiter gesellschaftlicher und wissenschaftlicher Diskurs behandelt das Thema Reichtum und Armut. Dabei wird häufig kritisiert, eine sehr kleine Bevölkerungsgruppe halte einen großen Anteil des gesellschaftlichen Vermögens, dieser Anteil werde weiter ausgebaut und niedrig besteuert an die Folgegenerationen übertragen. Dies entspricht nicht dem Leitbild einer meritokratischen Gesellschaftsordnung, bei der sich die Position in der Gesellschaft an „Leistung“ orientiert und die Mitglieder der Gesellschaft im Idealfall eine „verdiente“ Position einnehmen.

Entgegen der Breite des Diskurses ist die empirische Datenlage zur tatsächlichen Vermögenskonzentration in vielen Ländern sehr eingeschränkt. Dies gilt auch für Deutschland. Hier weisen alle großen Befragungen eine eklatante Datenlücke im Bereich jenseits niedriger einstelliger Millionenvermögen auf.

In der vorliegenden Studie wurden die Ergebnisse einer Machbarkeitsstudie vorgestellt, in der ein innovatives Konzept getestet wurde, um diese Datenlücke zu schließen. Alle Ergebnisse deuten klar darauf hin, dass die Umsetzung des Konzepts – Top-Vermögende über ihre Unternehmensbeteiligungen zu erfassen – im Rahmen einer größer angelegten Hauptuntersuchung erstmals für Deutschland eine umfangreiche Befragung von Top-Vermögenden erlaubt:

1. Ein wichtiger Vorteil des vorgeschlagenen Konzepts besteht darin, dass die Population der Top-Vermögenden mit den gleichen Instrumenten wie die Basispopulation befragt wird. Damit sind direkte Verglei-

che mit bereits vorhandenen SOEP-Daten zur Bevölkerung insgesamt möglich.

2. Die beiden verwendeten alternativen Ansprachen (reguläres SOEP-Informationsschreiben vs. Ansprache als Unternehmer) hatte keinen nennenswerten Effekt auf die Teilnahmebereitschaft.
3. Basierend auf 2.000 Adressen wurden 124 Personen erfolgreich befragt. Dies entspricht einer unbereinigten Teilnahmequote von 6%, wobei aber zu beachten ist, dass die Feldphase vorzeitig beendet wurde. Die Teilnahmebereitschaft unter denjenigen Personen, die kontaktiert werden konnten, lag bei ca. 23 Prozent.
4. Die Befragten sind reich: 56% der 124 Pretest-Befragten haben individuelle Nettovermögen in Millionenhöhe, rund 20% verfügen über ein Vermögen von mindestens 10 Mio. Euro, und sogar 5% über 50 Mio. Euro. Zudem wurde mindestens eine Person befragt im Bereich der 500 reichsten Deutschen laut Manager-Magazin.

Das hier präsentierte und getestete Konzept hat im Vergleich zu einer stratifizierten Stichprobenziehung etwa über die Höhe des regionalen Steueraufkommens (vgl. Private Haushalte und Finanzen) zwei wesentliche Vorteile: Erstens ist es deutlich zielgenauer hinsichtlich der Abgrenzung der Population der Top-Vermögenden. Dies senkt die Fallkosten je erfolgreich befragter vermögender Person. Zweitens lassen die Ergebnisse des Pretests erwarten, dass es im Rahmen einer Hauptuntersuchung erstmals gelingen dürfte, den oberen Bereich der Datenlücke in Deutschland zu schließen. Ein

entscheidender Vorteil des Konzepts im Vergleich zu Convenience Samples wie der HViD-Studie liegt darin, dass aufgrund der randomisierten Ziehung Inferenz-Statistik möglich ist und damit erstmals generalisierbare Aussagen für die Zielgruppe der Top-Vermögenden in Deutschland möglich sind.

Eine groß angelegte Befragung im Rahmen einer Hauptuntersuchung könnte wertvolle Daten zur Beschreibung von Vermögenskonzentration, -genese und -übertragung, sowie der Soziodemografie, Persönlichkeitsmerkmale und Aktivitäten (inkl. gesellschaftliches Engagement und Spendenverhalten) von „Reichen“ liefern. Darüber hinaus könnten die Daten wertvolle Erkenntnisse über die Validität der in vielen nationalen und internationalen Studien getroffenen Annahme liefern, wonach der obere Teil der Vermögensverteilung Pareto-verteilt sei. Eine Hauptuntersuchung in Form eines Panels kann perspektivisch wichtige Einsichten zur intra- und intergenerationalen Übertragung von Vermögen, sowie der Rolle von Steuern oder externer Schocks (z.B. in Form von Kapitalmarktschwankungen) auf die Vermögenskonzentration liefern.

7 Ausblick und Empfehlungen

Folgende Aspekte sollten vorbereitend für eine eventuelle Hauptuntersuchung beachtet werden:

1. Aufgrund der zeitlichen Vorgaben und enormen Herausforderungen beim Aufbau der Personendatenbank haben wir ein vergleichsweise einfaches Modell für die Imputation fehlender Umsatzinformationen in der Orbis-Datenbank verwendet, die Grundlage der Bewertung der

kumulierten Beteiligungen in der Personendatenbank und damit der stratifizierten Stichprobenziehung sind. Wenngleich die Ergebnisse des Pretests zeigen, dass es auch mit diesem einfachen Modell gelungen ist, die Grundgesamtheit erfolgreich entlang der Vermögensverteilung zu stratifizieren, sehen wir ein nennenswertes Potenzial zur Verbesserung des Imputationsmodells, das vor einer Hauptuntersuchung realisiert werden sollte.

2. Der Pretest hat gezeigt, dass die Adressqualität zufriedenstellend ist. Dennoch waren Teile des Adressmaterials unbrauchbar - weil Adressen unvollständig oder veraltet waren, oder es sich nicht um Privat- sondern um Unternehmensadressen handelte. Im Rahmen einer Hauptuntersuchung sollten daher weitere Arbeiten an der Qualität des Adressmaterials der Personendatenbank z.B. über Adressmelderecherchen erfolgen.
3. Die Teilnahmequote liegt je nach Abgrenzung zwischen 6% und 23%. Dies kann als Intervall für eine mögliche Hauptuntersuchung angesehen werden. Die Quote dürfte vermutlich unterhalb von 23% liegen, weil (a) Kantar sehr erfahrene Interviewer im Pretest eingesetzt hat; (b) Regionen mit kurzen Wegen und hoher Dichte von AnteilseignerInnen ausgewählt wurden; (c) der händische Aufwand bei den Adressmelderecherchen in einer großen Hauptuntersuchung weit höher ist als in einem Pretest mit kleiner Fallzahl.
4. Auch das SOEP-Vermögensmodul lässt sich in Hinblick auf seinen Einsatz unter Vermögenden optimieren: Im Sinne einer vereinfachten Ein-

gabe hoher nominaler Vermögenswerte (lange Ziffernfolgen) und einer detaillierteren Abfrage der vielfältigen Formen von Unternehmensbeteiligungen.

5. Teilweise werden hohe Vermögen in (Familien-)Stiftungen ausgelagert. Dies gilt z.B. für die Familie Albrecht, weshalb deren Beteiligungen in der Personendatenbank nicht so hoch wie auf den ersten Blick zu erwarten wäre. Vor einer Hauptuntersuchung könnte daher überprüft werden, ob Stiftungsregister für eine bessere Approximation der tatsächlichen Unternehmenswerte im Besitz dieses Personenkreises genutzt werden können.

Literatur

- Christian Westermeier and Markus M. Grabka. Große statistische Unsicherheit beim Anteil der Top-Vermögenden in Deutschland. *DIW Wochenbericht*, 82(7):123–133, 2015. URL <https://ideas.repec.org/a/diw/diwwob/82-7-3.html>.
- M. Ströing, M. M. Grabka, and W. Lauterbach. Hochvermögende in Deutschland. *DIW-Wochenbericht Nr.42/2016*, 2016.
- Sebnem Kalemli-Ozcan, Bent Sorensen, Carolina Villegas-Sanchez, Vadym Volosocyh, and Sevcen Yesiltas. How to construct nationally representative firm level data from the orbis database. *NBER Working Paper Series*, (21558), 2015.
- Joshua D. Coval and Tobias J. Moskowitz. Home Bias at Home: Local Equity Preference in Domestic Portfolios. *Journal of Finance*, 54(6):2045–2073, December 1999. URL <https://ideas.repec.org/a/bla/jfinan/v54y1999i6p2045-2073.html>.
- Martin Kroh, Simon Kuehne, Jan Goebel, and Friedrike Preu. The 2013 iab-soep migration sample (m1): Sampling design and weighting adjustment. *SOEP Survey Papers, Series C, Data Documentation*, (271), 2015.
- Joachim R. Frick, Markus Grabka, and Jan Marcus. Editing und multiple imputation der vermögensinformation 2002 und 2007 im soep. *SOEP Survey Paper*, (146), 12 2010.

Patrick Royston and Ian White. Multiple imputation by chained equations (mice): Implementation in stata. *Journal of Statistical Software, Articles*, 45(4):1–20, 2011. ISSN 1548-7660. doi: 10.18637/jss.v045.i04. URL <https://www.jstatsoft.org/v045/i04>.