Ferschli, Benjamin; Kapeller, Jakob; Schütz, Bernhard; Wildauer, Rafael

Working Paper

Bestände und Konzentration privater Vermögen in Österreich: Simulation, Korrektur und Besteuerung

Working Paper, No. 1805

Provided in Cooperation with:
Johannes Kepler University of Linz, Department of Economics

Suggested Citation: Ferschli, Benjamin; Kapeller, Jakob; Schütz, Bernhard; Wildauer, Rafael (2018) : Bestände und Konzentration privater Vermögen in Österreich: Simulation, Korrektur und Besteuerung, Working Paper, No. 1805, Johannes Kepler University of Linz, Department of Economics, Linz

This Version is available at:
https://hdl.handle.net/10419/183264

JOHANNES KEPLER UNIVERSITY LINZ

DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

Bestände und Konzentration privater Vermögen in Österreich

Simulation, Korrektur und Besteuerung

by

Benjamin FERSCHLI
Jakob KAPELLER
Bernhard SCHÜTZ
Rafael WILDAUER

Working Paper No. 1805
May 2018

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

bernhard.schuetz@jku.at

Bestände und Konzentration privater Vermögen in Österreich[⋆]

Simulation, Korrektur und Besteuerung

Benjamin Ferschli[*], Jakob Kapeller[**], Bernhard Schütz[**], Rafael Wildauer[***]

September/November 2017

⋆ Der spezielle Dank der Autoren gilt Stefan Steinerberger für seine wertvollen Hinweise im Rahmen der Erstellung dieser Arbeit.

[*] Johannes Kepler Universität Linz, Forschungsinstitut für die Gesamtanalyse der Wirtschaft.
[**] Johannes Kepler Universität Linz, Forschungsinstitut für die Gesamtanalyse der Wirtschaft und Institut für Volkswirtschaftslehre.
[***] University of Greenwich. Department of International Business and Economics and Greenwich Political Economy Research Centre.

Zusammenfassung

Dieser Bericht beschäftigt sich mit der Schätzung der Vermögensbestände privater Haushalte in Österreich. Basis und Anlass dieser Arbeit ist die Veröffentlichung der zweiten Welle des Household Finance Consumption Surveys (HFCS) des Europäischen Zentralbankensystems. Der HFCS bietet die gegenwärtig umfassendste Datengrundlage zur empirischen Auseinandersetzung mit privaten Vermögensbeständen in Europa und Österreich und leistet damit einen wesentlichen Beitrag zur ökonomischen Forschung.

Der vorliegende Bericht analysiert die Eigenschaften der Spitze der Vermögensverteilung und entwickelt auf dieser Basis ein Schätzverfahren für die vorliegenden Daten des HFCS in Österreich. Denn obwohl die Daten des HFCS für eine seriöse Auseinandersetzung mit bestehenden Privatvermögen unerlässlich sind, teilen sie doch die Schwierigkeiten der meisten umfragebasierten Vermögensdaten. Diese bestehen im Wesentlichen in der unzureichenden Erfassung der obersten Vermögensbestände einer Gesellschaft. Vermögensschätzungen, die auf derart unzureichenden Daten basieren liefern dabei notgedrungen verzerrte (*median-biased*) Ergebnisse, die sich dadurch auszeichnen, dass sie den Bestand und die Ungleichverteilung der Vermögen systematisch unterschätzen.

Ausgangspunkt für die vorliegende Untersuchung ist eine von Eckerstorfer et al. (2013, 2016) entwickelte Methode zur Schätzung des am oberen Verteilungsrand fehlenden Vermögens unter der Annahme einer Pareto-Verteilung (*non-observation bias*). Diese Methode wird dabei im Folgenden um den Aspekt selektiver Antwortverweigerungen *(non-response bias)* erweitert, da die Bedeutung von solcherart selektiver Antwortverweigerungen im Zuge der Erhebung zur zweiten Welle des HFCS angestiegen ist (OeNB, 2016a: 4-5, 2016b: 87). Zu diesem Zweck wird in einem ersten Schritt die statistische Eignung unterschiedlicher Schätzvarianten mittels Monte-Carlo Simulationen untersucht. Die Schätzung der Vermögen an der Spitze der Verteilung wird darauf aufbauend in einem zweiten Schritt durchgeführt.

Zusammenfassend ergibt die Vermögensschätzung unter der Annahme einer pareto-verteilten Vermögensspitze, dass sowohl der Bestand als auch die Ungleichheit des Vermögens in Österreich durch den HFCS unterschätzt wird. So steigt das geschätzte Gesamtvermögen der privaten Haushalte in Österreich um 319 Mrd. Euro (von 998 Mrd. auf 1317 Mrd. Euro), das Durchschnittsvermögen wächst um 83.000 Euro (von ca. 258.000 auf rund 341.000 Euro) und der Anteil des obersten Prozents am Gesamtvermögen steigt von 25% auf 41%. In einer vergleichbaren Studie basierend auf Daten der ersten Welle des HFCS aus dem Jahr 2010 (Eckerstorfer et al. 2013) führte eine solche Schätzung zu einem Anstieg des Gesamtvermögens um 249 Mrd. (von 1000 Mrd. auf 1249 Mrd. Euro), das Durchschnittsvermögen stieg um 67.000 Euro (von rund 266.000 auf etwa 333.000 Euro) und der Anteil des obersten Prozents erhöhte sich von 23% auf 37%.

Abschließend wird eine Schätzung des Aufkommenspotentials unterschiedlicher Vermögenssteuer-modelle durchgeführt. Diese Aufkommensschätzungen werden auf Basis der HFCS Daten sowie der mit Hilfe der Paretoverteilung geschätzten Vermögen durchgeführt; dabei werden auch mögliche Ausweicheffekte berücksichtigt. Die unter der Annahme einer pareto-verteilten Vermögensspitze berechneten Aufkommen schwanken je nach Steuertarif und den verwendeten Annahmen zum Ausweichverhalten zwischen 2,9 Mrd. und 8,3 Mrd. Euro.

Inhaltsüberblick

Anhang

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung und Forschungsfrage

Ein präziser Wissenstand über die Bestände und Verteilung der privaten Vermögen einer Volkswirtschaft ist nicht nur Ausgangspunkt wirtschaftspolitischer Richtungsentscheidungen, sondern leistet auch einen wesentlichen Beitrag zum wissenschaftlichen und öffentlichen Diskurs, der so transparenter und objektiver geführt werden kann. Aus akademischer Sicht ist derartiges Wissen zentral für die Beurteilung und Diskussion unterschiedlicher ökonomischer Theorien und Modelle, während in praktischer Hinsicht pragmatische Fragen nach den sozialen Folgen der Vermögensungleichheit oder möglichen Steueraufkommen relevant erscheinen. Besondere Bedeutung gewinnt diese Fragestellung vor dem Hintergrund sukzessiver steigender Ungleichheit im Bereich der Vermögen und Einkommen und ihren sozialen Folgen, wobei die jüngere Forschung vor allem die negativen Effekte zunehmender Ungleichverteilung von Vermögenswerten betont (siehe Guttmann/Plihon 2010, Stiglitz 2012, Piketty 2014). Vor diesem Hintergrund haben Fragen der Verteilung auch in der internationalen Debatte massiv an Bedeutung gewonnen: So konstatierte jüngst der Internationale Währungsfonds, dass zunehmende Ungleichheit mit negativen Folgen einhergeht.

> „*Widening inequality […] has significant implications for growth and macroeconomic stability, it can concentrate political and decision making power in the hands of a few, lead to a suboptimal use of human resources, cause investment-reducing political and economic instability, and raise crisis risk*." (IMF 2015: 5).

Trotz der großen praktischen Relevanz von Verteilungsdaten war das akademische und statistische Interesse an deren Erhebung lange relativ schwach ausgeprägt, sodass bis vor wenigen Jahren kaum verlässliche Daten zur Vermögensverteilung in entwickelten Ländern aufzufinden waren. Der Arbeit einzelner Forscher zum Trotz (Milanovic, 2005, 2016; Atkinson, 1995, 2008, 2014) hat das Interesse an Erhebungen zur Vermögensverteilung erst in den letzten Jahren – wesentlich bedingt durch die Finanzkrise und ihre Folgen – zugenommen. Die Gründe für eine solche verstärkte Aufmerksamkeit gehen dabei über die Rolle von Vermögensungleichheit als eine Ursache der Finanzkrise hinaus und betreffen auch allgemeinere Fragen, etwa jene nach dem Zusammenhang zwischen dem Ausmaß und der Verteilung privater Vermögensbestände und der allgemeinen wirtschaftlichen Entwicklung. Für die konkrete empirische Befassung mit privaten Vermögensbeständen und deren Verteilung gilt dabei natürlich, dass unser Verständnis aktueller ökonomischer Konstellationen immer nur so gut sein kann, wie die uns zur Verfügung stehenden Daten. An diesem Punkt setzt daher auch die vorliegende Studie zu den Beständen und der Verteilung der privaten Vermögen in Österreich an.

Trotz zahlreicher Verbesserungen in den letzten Jahren besteht auf der Ebene der Daten zur Vermögensverteilung weiterhin das Problem fehlender vollständiger Transparenzmechanismen auf nationaler und internationaler Ebene. So existiert zwar eine Zahl unterschiedlicher Datensätze, basierend auf unterschiedlichen Erhebungsgrößen und -Methoden, es fehlt aber weiterhin an einer vollständig normierten, harmonisierten und international vergleichbaren Datengrundlage. Die wichtigsten Anlaufstellen zu Vermögensdaten sind *Eurostat*, *nationale Statistikinstitute*, die *Luxembourg Wealth Study Database*, die *World Wealth and Inequality Database,* der *Survey of Consumer Finances (SCF)* und der *Household Finance and Consumption Survey* (HFCS) des Europäischen Zentralbankensystems. Gerade letztere Erhebung leistet einen wichtigen Beitrag zur Objektivierung und besseren Vergleichbarkeit privater Vermögen im europäischen Wirtschaftsraum. Der von der Europäischen Kommission initiierte HFCS wurde erstmals im Jahre 2013 publiziert und stellt die bislang umfassendste Erhebung von Vermögen im europäischen Raum dar. Die teilweise Harmonisierung der Erhebung, die von den teilnehmenden Nationalbanken durchgeführt wurde, erlaubt es die Daten für internationale Vergleiche heranzuziehen. Darüber hinaus bietet der HFCS für die meisten Mitgliedsländer die einzige bzw. umfassendste Datenbasis zur Analyse privater Vermögen. Eine zweite Welle

des HFCS, die 2016 veröffentlicht wurde, schließt an die Zielsetzung der ersten Welle an und beinhaltet Verbesserungen der Erhebungsmethoden sowie eine Erweiterung der teilnehmenden Länder.[1]

Obwohl der Household Finance Consumption Survey (HFCS) eine vergleichsweise solide Basis für die Schätzung von Vermögensbeständen und deren Verteilung bildet, ist auch der HFCS von den *generellen statistischen Problemen* umfragebasierter Vermögensschätzungen betroffen. Diese bestehen vor allem in einer unzureichenden Erfassung der Spitze der Vermögensverteilung sowie in selektiven Antwortverweigerungen. Generell gilt, dass Vermögensverteilungen einen „*fat tail*" aufweisen, also dass (besonders) reiche Haushalte zwar nur sehr selten vorkommen (und daher zumeist nicht Teil der Zufallsstichprobe sind), aber einen großen Einfluss auf die finalen Schätzwerte nehmen. Diese Problematik der Erfassung der Haushalte am oberen Rand der Verteilung (*non-observation* bias[2]) führt dabei in den meisten Fällen zu einer Unterschätzung der Größe und Ungleichverteilung der tatsächlichen Vermögensbestände (Avery et al. 1986, Kennickell 2005, Eckerstorfer et al. 2016 bzw. Kapitel 4 der vorliegenden Arbeit). Eine zweite Quelle von Unsicherheit in den HFCS-Daten bildet die Möglichkeit selektiver Antwortverweigerungen (*non-response bias*) – also, dass reichere Haushalte eher dazu tendieren eine Teilnahme an der Befragung zu verweigern (D'Alessio and Faiella 2002, Kennickell 2008, Kennickell and McManus 1993, OeNB 2016a, Osier 2016, Singer 2006). Theoretisch können beide Formen der Verzerrung durch ein gezieltes „oversampling" besonders vermögender Haushalte kompensiert werden – allerdings wurde im Zuge der Erhebung der österreichischen HFCS-Daten kein derartiges Oversampling-Verfahren eingesetzt.[3] Diese beiden Einflussfaktoren stellen daher ein Problem für die Verwendung der Daten der HFCS zur Bestimmung der Bestände und Verteilung der österreichischen Privatvermögen dar. Die vorliegende Untersuchung versucht diese Schwächen mit Hilfe eines geeigneten statistischen Verfahrens zumindest teilweise zu kompensieren.

Die vorliegende Arbeit zielt darauf ab eine realistischere Schätzung des Vermögens der privaten Haushalte zu erzielen, wobei angenommen wird, dass der oberste Rand (i.e. die Verteilung der reichsten Vermögen) einer Pareto-Verteilung folgt. Die Arbeit baut auf der von Eckerstorfer et al. (2013, 2016) vorgeschlagenen Methode auf und entwickelt diese weiter um auch das Problem des non-response bias zu berücksichtigen, das in Eckerstorfer et al. (2013, 2016) außer Acht gelassen wurde.

Dieser Forschungsbericht ist wie folgt aufgebaut: Im nachfolgenden Kapitel 2 wird die Anwendbarkeit einer Pareto-Verteilung auf Vermögensdaten besprochen und die Eigenschaften des HFCS für Österreich genauer dargestellt. Anschließend werden typische Probleme der statistischen Erfassung von Vermögen diskutiert sowie Methoden zu deren Behebung vorgestellt. In Kapitel 3 werden die Daten für Österreich deskriptiv aufgearbeitet, sowie Unterschiede zwischen den beiden Wellen des HFCS analysiert. Kapitel 4 untersucht die statistische Eignung unterschiedlicher Schätzverfahren der Verteilungsspitze mit Hilfe von Monte-Carlo Simulationen um das unter gegebenen Umständen bestgeeignetste Verfahren zu bestimmen. In Kapitel 5 wird schließlich eine pareto-basierte Schätzung des Gesamtvermögens präsentiert. Kapitel 6 diskutiert die Robustheit der so gewonnen neuen Ergebnisse. Kapitel 7 präsentiert Ergebnisse einer Vermögenssteueraufkommensschätzung auf Basis der entwickelten Methode. Kapitel 8 fasst schließlich die gewonnenen Ergebnisse nochmals zusammen.

[1] Teilnehmende Länder in Welle I waren: Österreich, Belgien, Zypern, Deutschland, Spanien, Finnland, Frankreich, Griechenland, Italien, Luxemburg, Malta, Niederlande, Portugal, Slowenien und die Slowakei. Welle II wurde um Daten für Estland, Ungarn, Irland, Lettland und Polen erweitert.

[2] Der Begriff „non-observation bias" wird im Rest der vorliegenden Studie verwendet um zu verdeutlichen, dass herkömmliche Verfahren zur Schätzung des Haushaltsvermögens „median-biased" sind. Dies bedeutet, dass im Fall einer mehrfachen Wiederholung der Untersuchung der Median der geschätzten Gesamtvermögenswerte unterhalb des wahren Wertes liegt (siehe dazu auch die Monte-Carlo Simulationen in Kapitel 4).

[3] Alle teilnehmenden Länder des HFCS, mit Ausnahme von Griechenland, Österreich und Malta, wenden hingegen, in verschieden starkem Ausmaß, Oversampling an.

2. Methodisches Vorgehen

Die meisten konventionellen statistischen Verfahren – darunter auch das Verfahren der (stratifizierten) Zufallsziehung und Randomisierung, das dem HFCS zu Grunde liegt – gehen von Variablen als Standardfall aus, die so verteilt sind, dass die Abweichungen zwischen geschätztem und wahren Wert grob einer Normalverteilung folgen. Eine allgemeine Herausforderung der Sozialstatistik ist dabei, dass viele Variablen von gesellschaftlicher Relevanz einer solchen Annahme nicht genügen, sondern davon abweichende statistische Eigenschaften aufweisen. Eine im sozialen Raum häufig auftretende Verteilung ist dabei die Potenzverteilung (auch: Zipfsche Verteilung, in der Ökonomie und im Folgenden als „Pareto-Verteilung“ bezeichnet), die etwa geeignet ist Worthäufigkeiten in Texten, die Größenverteilung von Städten oder die Verteilung der Anzahl verkaufter Bestseller statistisch abzubilden (Newman 2005, Gabaix 2016). Der Wert der Pareto-Verteilung liegt dabei vor allem darin, dass diese im Stande ist, den oberen Rand der relevanten Verteilungen statistisch zu beschreiben – ein Umstand, der insofern von Interesse ist, als Messwerte an der Spitze der jeweiligen Skala einen besonders großen Einfluss auf die finalen Schätzwerte nehmen. Genau diese Eigenschaft – und der Umstand, dass Vermögen in entwickelten Gesellschaften zumeist einer solchen Pareto-Verteilung folgen – macht die Pareto-Verteilung auch für die ökonomische Verteilungsforschung zu einem besonders wertvollen Instrument.

Zur genaueren Illustration zeigt Abbildung 1 eine normalverteilte Variable (links oben) im Vergleich zu drei Variablen, deren oberer Rand einer Potenzverteilung folgt, wobei die Messwerte jeweils vom größten zum kleinsten Wert sortiert sind. Die Normalverteilung zeichnet sich dadurch aus, dass die Werte einer normalverteilten Variable relativ stark rund um den Mittelwert konzentriert liegen. Ausreißer sind sehr unwahrscheinlich und Rückschlüsse auf die Population sind schon mit relativ kleinen Zufallsstichproben möglich.

An Abbildung 1 lassen sich daher einige wesentliche Unterschiede zwischen normalverteilten Variablen und Variablen, die an der Spitze einer Potenzverteilung folgen festmachen: Erstens, zeigen normalverteilte Variablen relativ kontinuierliche Steigerungsraten, während potenzverteilte Variablen starke Diskontinuitäten aufweisen. Zweitens liegen Mittelwert und Median bei annähernd normalverteilten Variablen üblicherweise nah beieinander, während bei potenzverteilten Größen der Mittelwert den Median zumeist um ein Vielfaches übersteigt. Die Relation von Mittelwert und Median kann dabei auch als ein erstes grobes Verteilungsmaß herangezogen werden. Drittens, ist die Streuung der Messwerte bei Potenzverteilungen typischerweise weitaus größer, ein Effekt, der wesentlich durch den oberen Rand der Verteilung bestimmt ist.

Eine der möglichen theoretischen Ursachen für die Häufung solcher Potenzverteilungen im sozialen Raum liegt dabei in der Theorie kumulativer Effekte, nach der bestehende Stärken (bestehende Vermögen, bestehendes akademisches Prestige, bestehende urbane Vielfalt) auch höhere Wachstumsraten nach sich ziehen (Gibrat 1931, Rigney 2010) – ein Ansatz, der auch für die Theorie der Vermögensbildung relevant erscheint (Borgherhoff-Mulder et al. 2009), in der vorliegenden Arbeit aber nicht weiter verfolgt werden kann.

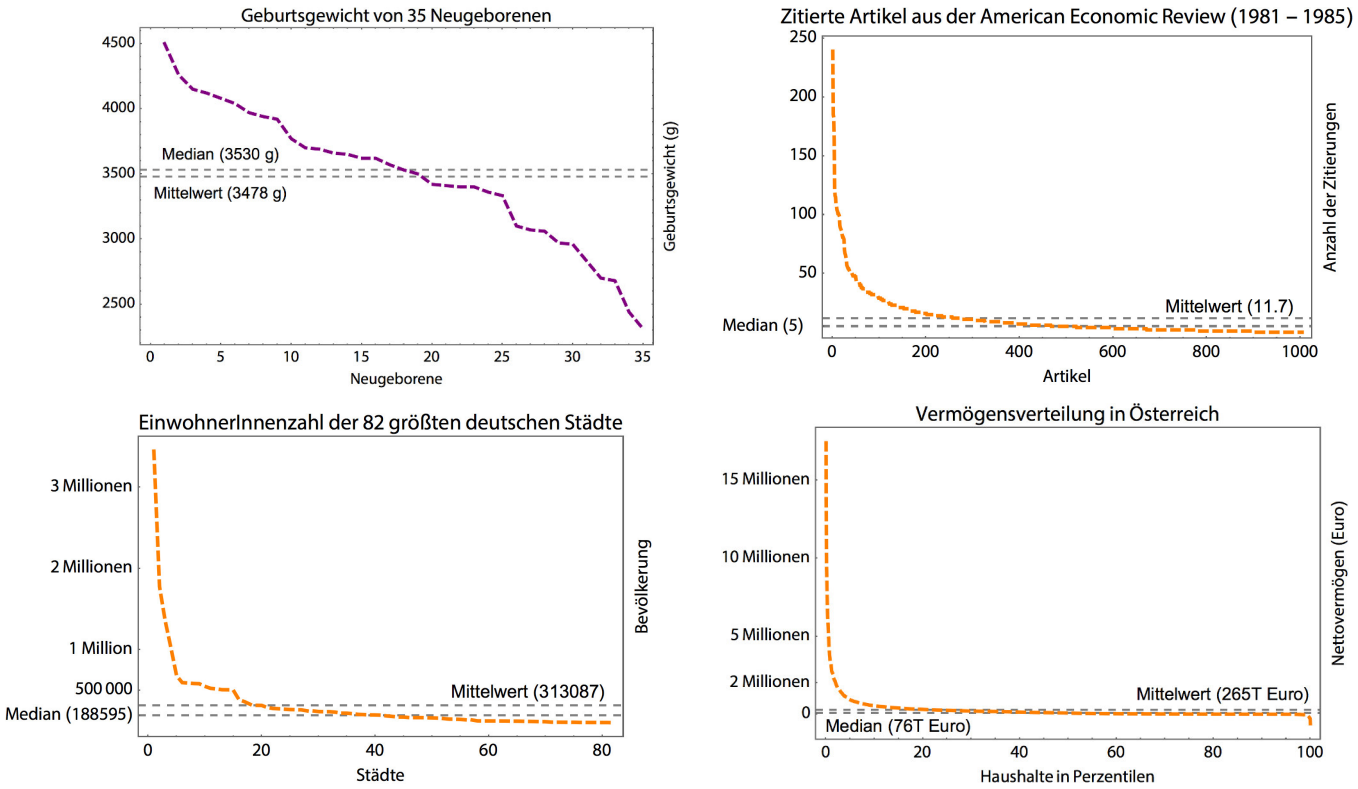


Abbildung 1: Vergleich einer normalverteilten Variable (violett) mit drei Variablen, deren oberer Rand einer Potenzverteilung folgt (orange).

Grundsätzlich beschreibt eine Pareto-Verteilung also ein Potenzgesetz das in vielen natur- und sozialwissenschaften Fragestellungen zur Anwendung kommt und auch für die Vermögensforschung von zentraler Relevanz ist. In den meisten empirischen Vermögensverteilungen zeigt sich eine entsprechende Konzentration von wenigen, hohen Vermögenswerten am oberen Rand, die mittels einer Pareto-Verteilung näherungsweise beschrieben werden können. Die Verteilungsfunktion letzterer sieht folgendermaßen aus:

$$F(x) = \Pr(X \leq x) = 1 - \left(\frac{m}{x}\right)^{\alpha} \quad \forall\, x \geq m \qquad (1)$$

Hier steht x für das gemessene Nettovermögen eines Haushaltes. Der Parameter m ist der Schwellenwert, ab dem die Verteilung der Nettovermögen einer Pareto-Verteilung folgt. Der Formparameter α (das sogenannte „Pareto Alpha“) beschreibt schließlich die genaue Form dieser Verteilung.

Die Annahme der Pareto-Verteilung des oberen Randes der Vermögen wird dabei im Folgenden aufrechterhalten, da die Pareto-Verteilung nicht nur denselben Erkenntnisgewinn wie konkurrierende Verteilungsannahmen (wie etwa die Dagum oder Singh-Madalla) bringt, sondern darüber hinaus mit weniger Annahmen einhergeht, robustere Ergebnisse liefert und daher im wissenschaftlichen Diskurs stärker etabliert ist. Jüngere Studien, die eine Pareto-Verteilung zur Schätzung des obersten Vermögenssegments herangezogen haben, inkludieren unter anderem: Atkinson (2006), Cowell (2009) und Klass et al. (2006).

2.1. Datenquelle und Erhebung

Datengrundlage der vorliegenden Arbeit ist die mit Dezember 2016 veröffentliche zweite Welle des Household Finance and Consumption Survey (HFCS) der ECB. Beide Wellen folgen einem international standardisierten Erhebungsverfahren (ECB 2016a). Im Fall von Österreich hat sich der Erhebungszeitraum über neun Monate, von Juni 2014 bis Februar 2015, erstreckt (vgl. Welle I: 09/10 – 05/11). Beide Wellen wurden durch die Österreichische Nationalbank (OeNB) unter Mitarbeit des Instituts für empirische Sozialforschung (IFES) erhoben. Die Stichprobe für Österreich umfasst 2.997 Beobachtungen, im Vergleich zu 2.380 aus Welle I, und repräsentiert rund 3,9 Millionen Haushalte. Eine Beobachtung aus dieser Stichprobe entspricht einem Haushalt, dessen Vermögenswerte von einer KompetenzträgerIn[4] im Laufe eines Interviews beschrieben worden ist. Die „Repräsentativität" der so erhobenen Haushaltswerte für die Gesamtpopulation wird durch die von der OeNB zugewiesene Gewichtungen dargestellt. Methodisch relevant ist hier noch die Verwendung von multipler Imputation in der Datenerhebung durch die OeNB. Dies bedeutet schlichthin, dass fehlende Datenpunkte aus dem Zusammenhang anderer Datenpunkte geschätzt werden. Diese Methode hilft dabei einzelne ausgelassene Werte (item non-response) zu ergänzen, erfordert aber die Generierung mehrerer, ähnlicher Datensätze, so genannte Imputationen, um die im Schätzverfahren auftretenden statistischen Unsicherheiten adäquat abzubilden. Für die Generierung von Schätzwerten müssen dabei alle Imputationen berücksichtigt werden (Rubin 1987, Little und Rubin 2002). Der publizierte Datensatz zur zweiten Welle des HFCS umfasst fünf derartige Imputationen. Strukturelle Probleme der umfragebasierten Vermögensschätzung, wie non-response oder non-observation, werden durch das Imputationsverfahren *nicht* korrigiert. In der nachfolgenden Tabelle sind ausgewählte Erhebungsindikatoren der HFCS Wellen I und II für Österreich zum Vergleich dargestellt.

Tabelle 1: Eckdaten zur HFCS-Erhebung in Österreich – Welle I und Welle II

HFCS-Erhebung	Netto-stichprobe	repräsentierte Haushalte	Antwortrate	Veweigerungs-rate	Oversampling-rate der top 10%
Welle I	2.380	3.773.956	55,7%	39,6%	1
Welle II	2.997	3.862.526	49,8%	44,1%	-7

Quelle: ECB (2016a: 35,39, 43, 2013: 38, 41, 45)

Während sich die Größe der Nettostichprobe, und die dadurch repräsentierten Haushalte, kaum geändert haben hat sich im Zuge der zweiten Wellen des HFCS die Antwortrate (*response-rate*), also das Verhältnis von Personen, die auf die Befragungsanfrage reagiert haben, zur Gesamtstichprobe, weiter *verringert*. Im Gegensatz dazu hat die Verweigerungsrate (*refusal-rate*), also das Verhältnis von Personen, die erfolgreich kontaktiert wurden, dann aber nicht an der Befragung teilgenommen bzw. es abgelehnt haben Auskunft über ihre Vermögen zu geben, zur Gesamtstichprobe zugelegt. Antwortverweigerungen sind dabei insofern problematisch, als dass diese nicht zufällig auftreten, sondern mit dem Reichtum eines Haushaltes zunehmen (D'Alessio and Faiella 2002, Kennickell 2008, Kennickell and McManus 1993, OeNB 2016a, Osier 2016, Singer 2006). Eine wesentliche Frage für die Beurteilung der Vergleichbarkeit beider Wellen des HFCS ist also, ob sich der Zusammenhang zwischen bestehendem Vermögen und der Wahrscheinlichkeit der Antwortverweigerung verändert hat – ein Aspekt, der sich aus der Verweigerungsrate allein nicht erschließen lässt.

[4] Als KompetenzträgerIn wird jene Person bezeichnet, die aus Sicht der Haushaltsmitglieder die beste Kenntnis über die Haushaltsfinanzen, also Verbindlichkeiten, Vermögen, Einkommen und Ausgaben des Haushalts, hat. Diese Person beantwortete alle Fragen, die sich auf den gesamten Haushalt beziehen.

Relevant ist in diesem Kontext vor allem die Veränderung der "*effective oversampling rate*" (EOR), die in Österreich aufgrund des fehlenden Oversamplings generell niedrig ausfällt (zum Vergleich, Deutschlands Oversampling rates für Welle I und II waren 117 bzw. 141). Diese gibt an, inwieweit reiche Haushalte stärker oder schwächer in der Stichprobe repräsentiert sind als ihr relativer Anteil in der Gesellschaft.[5] Der Rückgang der Oversampling-Rate zeigt dabei an, dass reichere Haushalte in der zweiten Welle des HFCS eine geringere Beteiligung aufweisen als in der ersten Welle – ein Umstand der einen stärkeren Zusammenhang zwischen bestehendem Vermögen und Antwortverweigerung nahelegt und der aus diesem Grund in Kapitel 3.2 genauer untersucht wird. Generell zeigt, eine negative EOR, wie für Österreich in Welle II, an, dass wohlhabende Haushalte unterproportional repräsentiert sind.

2.2. Statistische Verzerrungen im Kontext von Vermögensbefragungen

Trotz des bereits erwähnten präzedenzlosen Umfangs des HFCS leiden die so gesammelten Daten (für Länder ohne weitreichende oversampling Verfahren) de facto weiterhin an den zentralen methodischen Schwächen von Vermögensumfragen. Die zentralsten dieser Schwierigkeiten entstehen hierbei durch „*non-response*", „*non-observation" und* „*underreporting*". Unter *non-observation* ist das Problem zu verstehen, dass Zusammensetzung und Vermögen der reichsten Haushalte durch das standardisierte Zufallssampling der relevanten Umfragen, auf Grund der geringeren Zahl von reichen Haushalten, nicht adäquat erfasst werden und dadurch die Ungleichheit der Vermögen unterschätzt wird. „*Non-response*" hingegen weist auf jene Verzerrungen hin die entstehen, wenn Antwortbereitschaften sich in Zielpopulationen unterscheiden, etwa wenn reichere Haushalte öfter die Teilnahme an Vermögensumfragen ablehnen. Eine Strategie, die von vielen Nationalbanken verwendet wird, um speziell diesen ersten beiden Schwierigkeiten beizukommen, ist „oversampling". Dabei wird versucht, verhältnismäßig mehr „reiche" Haushalte in die Befragungsstichprobe aufzunehmen, also zu überrepräsentieren, und damit der eigentlichen Unterrepräsentation entgegenzuwirken. „*Underreporting"* hingegen bezieht sich ähnlich dazu auf die Unterschätzung von Vermögenswerten durch Umfrageteilnehmer.

Würden oben genannte Phänomene rein zufällig auftreten, wäre dies kein Problem. Wenn beispielsweise aber Finanzvermögen im Sinne von *underreporting* unterschätzt wird und Finanzvermögen überwiegend von wohlhabenden Haushalten gehalten wird, stellt dies ein zu Verzerrungen führendes Problem dar. Da es empirische Nachweise gibt, dass reichere Haushalte höhere non-response aufweisen (D'Alessio and Faiella 2002, Kennickell 2008, Kennickell and McManus 1993, OeNB 2016a, Osier 2016, Singer 2006) ist dies problematisch. Verringern sich die response-rates einer Umfrage bzw. verändert sich die Partizipation besonders vermögender Haushalte, wie für Österreich zwischen Welle I und II, dann bedeutet dies somit, dass sich die Anzahl der reichsten Haushalte in der Stichprobe verringert und sich daher die Qualität der Daten in dieser Hinsicht verschlechtert. Die im Folgenden angestellten Überlegungen beschäftigen sich dabei mit Verzerrungen, die aufgrund von *non-observation* und *non-response* entstehen, jedoch nicht mit jenen die auf *underreporting* beruhen.

[5] Wenn der Anteil reicher Haushalte des obersten Dezils in der Stichprobe etwa 10% ist, so beträgt die effective oversampling rate (of the top 10%) 0. Wenn der Anteil der Haushalte des wohlhabensten Dezils in der Stichprobe 20% entspricht, ist die EOR gleich 100. Sprich, es sind 100% mehr reiche Haushalte in der Stichprobe als wenn alle Haushalte gleich gewichtet wären.

2.3. Alternativen Strategien zur Schätzung von privaten Vermögen

Grundsätzlich bestehen mehrere Möglichkeiten zur Erhebung von Vermögensdaten. Der Rückgriff auf Steuerdaten ist einer davon. Da Vermögenssteuern nicht in jedem Land eingehoben werden, ist deren Verwendung aber nur für wenige europäische Länder möglich.[6] Ähnlich verhält es sich mit Erbschaftssteuerdaten. Dabei bestehen außerdem die Probleme, dass (a) nicht erfasst wird was von der Besteuerung ausgenommen ist, und (b) nicht erfasst wird, was schlicht hinterzogen wird. Eine andere Strategie besteht darin, aus Einkommenssteuerdaten, konkret auf Basis der angegebenen Kapitaleinkommen auf Vermögenswerte zu schließen. Die zentrale Komplikation dabei ist, dass die hierfür nötigen Kapitalertragsraten schlichtweg geschätzt werden müssen. Eine letzte Möglichkeit sind die bereits diskutierten, speziell konzipierten Vermögensumfragen. Die Vorteile dieser Erhebungsmethode, trotz der oben geschilderten statistischen Probleme, sind nach Vermeulen (2016), dass sie konzipiert sind, alle einzelnen Vermögenskomponenten zu erfassen und für alle Haushalte repräsentativ zu sein. Darüber hinaus ergeben sie zusätzlich zum Vermögensbestand des Haushalts eine Vielzahl and weiterer Charakteristika die eine weiterführende, detaillierte Analyse ermöglichen. Obwohl umfassende und transparente internationale Steuerdaten für den reinen Zweck der Haushaltsvermögensschätzung zu bevorzugen wären, bieten Vermögenserhebungen wie der HFCS oftmals die verhältnismäßig beste Basis für Vermögensschätzungen. Diese Feststellung motiviert klarer, warum Anstrengungen zur Lösung der Probleme von „non-response“, „non-observation“ und „underreporting“ von derartiger Wichtigkeit sind und es deshalb wert sind weiterverfolgt zu werden. Im Folgenden werden für diese Studie relevante Schätzmethoden genauer besprochen und angeführt.

Eine häufig verwendete Methode ist die Verwendung von Reichenlisten. Unter Reichenlisten sind die meist journalistisch zusammengetragenen Schätzungen der reichsten Haushalte/Personen in Volkswirtschaften zu verstehen. Die wohl bekannteste diesbezügliche Publikationsplattform ist das Magazin *Forbes*. Die Verwendung von Reichenlisten adressiert direkt das Problem der non-observation, also das Fehlen der reichsten Haushalte in der Stichprobe. Reichenlisten beinhalten aber ihre eigenen Schwierigkeiten. So ist deren Erhebung nicht frei von Problemen, da sie zum einen auf den Schätzungen von Journalisten basieren und zum anderen Vermögenswerte oft nicht sauber zwischen Individuum getrennt , sondern vielmehr ganzen Familien zugeordnet werden.

Ein anderes Beispiel ist die von Eckerstorfer et. al (2013, 2016) entwickelte Methode, zur (partiellen) Schätzung des aufgrund von non-observation nicht abgebildeten Vermögens. In Eckerstorfer et al. führt die Schätzung nicht erfasster Vermögen reicher Haushalte auf Basis dieser Methode zu einer bedeutend realistischeren Schätzung des Gesamtvermögens. Die Schritte des Schätzverfahrens lassen sich wie folgt zusammenfassen: zuerst werden auf Basis der verfügbaren HFCS Daten für zunehmend größere Abschnitte der geordneten Daten Pareto-Verteilungsfunktionen geschätzt (100.-70. Perzentil). Aus diesen Pareto-Verteilungen wird mit Hilfe einer Maxi-Min Analyse der Teststatistiken eines Cramer-von-Mises Tests (ein Test zur Überprüfung der Verteilung von Daten mit einer hypothetischen Verteilung) der Punkt gewählt, ab dem die zu Grunde liegenden Daten am ehesten einer Pareto-Verteilung entsprechen, woraufhin der Pareto-Steigungsparameter (Pareto-Alpha) an diesem Punkt bestimmt wird. In einem nächsten Schritt werden alle enthaltenen Haushalte ab einem gewissen Schwellenwert (Haushalte mit mehr als 4.000.000 Euro Nettovermögen) entfernt, da die dort vorgefundenen Vermögenswerte als nicht repräsentativ für das tatsächliche Vermögen eingeschätzt werden. In einem letzten Schritt werden synthetische Haushalte auf Basis der zuvor geschätzten Pareto-Verteilung generiert, dem Datensatz hinzugefügt und das Vermögen neuerlich geschätzt.

[6] In Österreich etwa wurde die Vermögenssteuer 1994 abgeschafft.

Eine ähnliche Methode auf Basis einer Pareto-Verteilung wurde von Vermeulen (2014) entwickelt. Zentraler Unterschied zur Methode von Eckerstorfer et al. ist, dass Vermeulen Beobachtungen aus Reichenlisten in seine Schätzung integriert. Eckerstorfer et al. verwenden hingegen Reichenlisten nur als Robustheitstest. Außerdem sind Vermeulens Reichenlisten allgemeiner gehalten (Forbes-Billionaires) als bei Eckerstorfer et al., die sich auf die umfassendere, auch Millionäre inkludierende Trend Publikation der reichsten ÖsterreicherInnen beziehen. Bedingt durch diese unterschiedliche Vorgehensweise unterscheidet sich auch die Art des Schätzers der für die Schätzung des Pareto-Alphas verwendet wird. Der von Eckerstorfer et al. verwendete Maximum-Likelihood-Schätzer steht hier einem OLS-Schätzer bei Vermeulen gegenüber. In seiner Arbeit aus 2016 inkludiert Vermeulen ferner einen Vorschlag zur Behebung von *underreporting* über einen „Korrekturfaktor“ der Schätzergebnisse mit den aggregierten Werten aus der Volkswirtschaftlichen Gesamtrechnung (VGR) gegenüberstellt. Großer Vorteil der Methode von Eckerstorfer et al. ist, dass diese eine statistische Basis für die Wahl des Startpunkts der Pareto-Verteilung der oberen Vermögenswerte verwendet (Rückgriff auf Cramer-von-Mises Test und Maxi-Min Analyse) und so die Schätzmethode weiter präzisiert.

3. Deskriptive Merkmale der untersuchten Daten

Um einen ersten Überblick über die im Rahmen der zweiten Welle des HFCS erhobenen Daten zu ermöglichen (siehe auch: OeNB 2016a), werden im Folgenden einige deskriptive Merkmale des untersuchten Datensatzes vorgestellt (3.1). In einem zweiten Schritt werden die im HFCS erfassten Beobachtungen am oberen Rand der Verteilung genauer analysiert. An dieser Stelle wird, vor dem Hintergrund der veränderten Kontextvariablen der Datenerhebung (vgl. Tabelle 1) und deren Relevanz für die Identifizierung geeigneter Schätzmethoden, auch ein expliziter Vergleich zu den Daten aus der ersten Welle des HFCS vorgenommen (3.2.).

3.1. Allgemeine deskriptive Statistiken

Das aus den Originaldaten des HFCS II für Österreich geschätzte Nettogesamtvermögen beträgt 998 Mrd. Euro. Zum Vergleich: das Gesamtvermögen der Welle I lag mit 1.000 Mrd. Euro knapp darüber; die österreichischen Vermögensbestände wären demnach zwischen 2010 und 2014 leicht zurückgegangen. Auch bei den durchschnittlichen Vermögenswerten ist ein entsprechender Rückgang zu verzeichnen – so sinkt das Durchschnittsvermögen von rund 265.000 Euro auf etwa 258.000 Euro ab. Beim Median der Vermögen hingegen weisen die Daten der zweiten Welle einen deutlichen Anstieg gegenüber der ersten Welle aus (von rund 80.000 Euro auf rund 90.000 Euro).

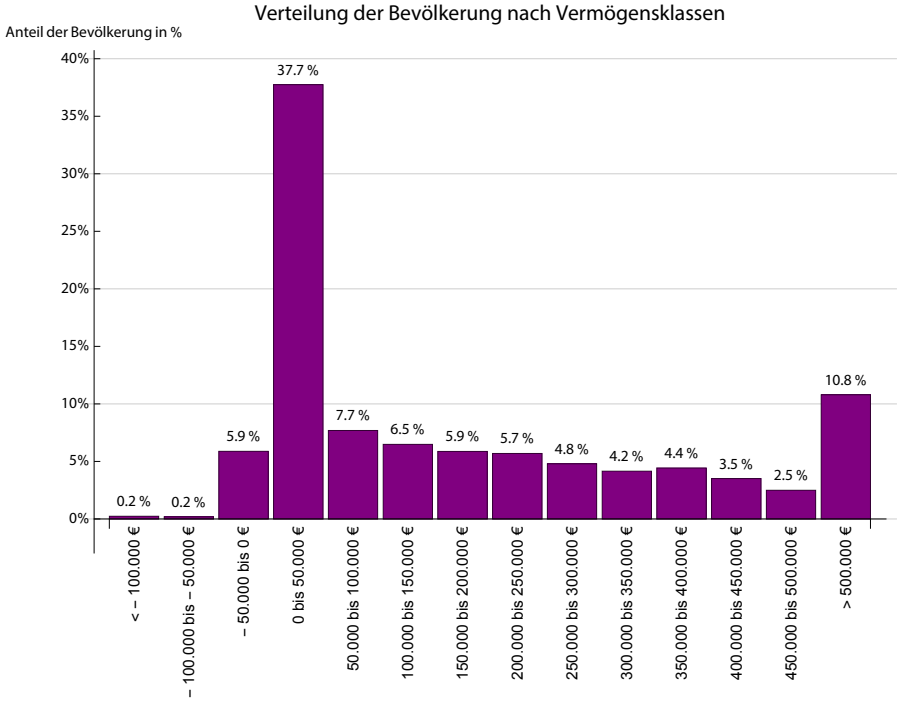


Abbildung 2: Vermögensverteilung in Österreich nach Vermögensklassen (Originaldaten HFCS II)

Eine erste mögliche Darstellung der in der zweiten Welle des HFCS gemessenen Vermögensverteilung bietet die Zuordnung der beobachteten Haushalte zu unterschiedlichen Vermögensklassen. Abbildung 2 nimmt eine solche Zuordnung vor und zeigt die Verteilung der österreichischen Bevölkerung auf 14 Vermögensklassen, wie sie aus der zweiten Welle des HFCS hervorgeht. Die Vermögensklassen beginnen bei -100.000 Euro, enden bei 500.000 Euro und umfassen eine Bandbreite von jeweils 50.000 Euro. Dabei ist zu sehen, dass sich mehr als ein Drittel der österreichischen Bevölkerung in der Vermögensklasse von 0 bis 50,000 Euro wiederfindet (knapp 38%). Während knapp mehr als 10% der Bevölkerung ein Nettovermögen von über 500.000 Euro aufweisen, sind knapp über 6% der

österreichischen Haushalte Netto-Schuldner, die offenen Schulden übersteigen in diesen Fällen also das verfügbare Vermögen.

Während die Darstellung nach Vermögensklassen einen guten Überblick über die Häufigkeitsverteilung der privaten Vermögenswerte innerhalb der Bevölkerung bietet, lässt sie keine Rückschlüsse auf die genaue Verteilung der Vermögen am oberen Rand der Verteilung zu, da dieser in Abbildung 2 nur pauschal ausgewiesen ist. Die nachstehende Tabelle 2 hingegen zeigt das Gesamt- und Durchschnittsvermögen der obersten 5 Perzentile der Vermögensverteilung im Detail und lässt erkennen, dass unter jenen 10.8% der Haushalte mit einem Nettovermögen größer als 500.000 Euro eine erhebliche Variation besteht. Speziell markant ist hierbei der Unterscheid zwischen dem 100. Perzentil und dem 99., mit einer Gesamtvermögensdifferenz von 192,2 Mrd. Euro.

Tabelle 2: Vermögensverteilung der obersten 5 Perzentile HFCS II (Originaldaten HFCS II)

Perzentil	Gesamtnettovermögen im Perzentil	Durchschnittsnettovermögen im Perzentil
96	€ 33 Mrd.	€ 847.449
97	€ 37,4 Mrd.	€ 980.399
98	€ 47,1 Mrd.	€ 1,22 Mio.
99	€ 62,4 Mrd.	€ 1,62 Mio.
100	€ 254,5 Mrd.	€ 6,7 Mio.

Nachstehend folgen zwei Darstellungen der kumulativen Verteilungsfunktion der österreichischen Vermögen. Eine kumulative Verteilungsfunktion zeigt, welcher Anteil der Beobachtungen unterhalb eines gewissen Vermögenswerts liegt. Grundsätzlich bildet diese also das Verhältnis von Vermögen relativ zur Bevölkerung ab und erlaubt Aussagen der Art: die oberen 20% der Vermögensverteilung verfügen über ein Nettovermögen von mehr als 337.000 Euro. Dabei zeigt Abbildung 3 eine kumulative Verteilungsfunktion anhand der absoluten Vermögenswerte; die Skalierung der y-Achse ergibt sich hier aus jener Beobachtung im Datensatz mit dem größten Vermögen.

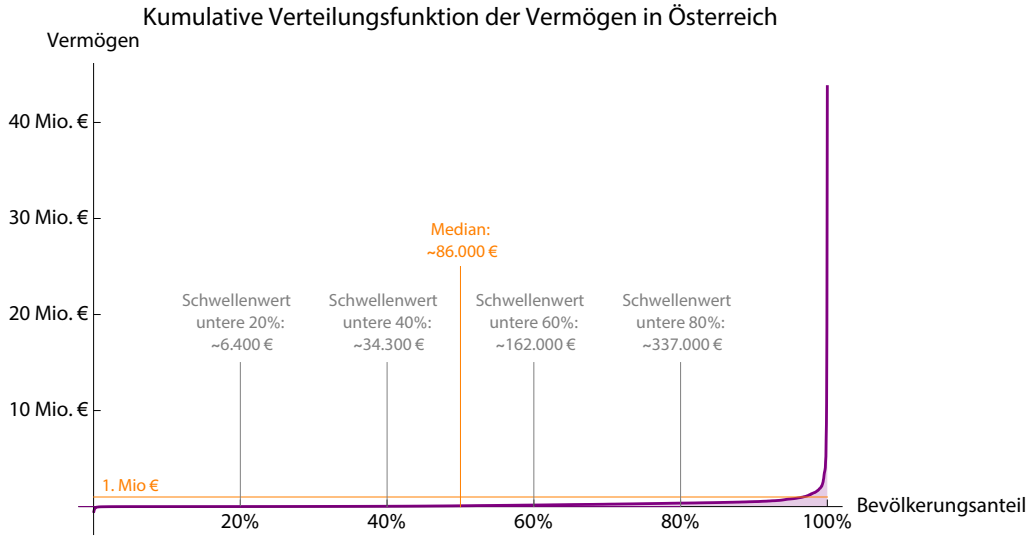


Abbildung 3: Kumulative Verteilungsfunktion der österreichischen Privatvermögen (Originaldaten HFCS II)

Abbildung 4 hingegen zeigt eine so genannte Lorenzkurze, in der die Skalierung der Vermögen normalisiert wird – die kumulative Verteilungsfunktion wird damit wesentlich übersichtlicher und leichter zu interpretieren; die Interpretation entspricht jener von Abbildung 3. Die Lorenzkurve bildet auch die Grundlage für die Berechnung des Gini-Koeffizienten, der oftmals als Maß zur Messung von Ungleichheit herangezogen wird. Der Gini-Koeffizient beschreibt den Unterschied zwischen einer hypothetischen Situation absoluter Gleichverteilung (graphisch repräsentiert durch die 45-Grad-Linie) und der realen Verteilungssituation. Die Fläche zwischen den beiden Linien repräsentiert das durch den Gini-Koeffizient gemessene Ausmaß der Ungleichheit – je größer der Wert, desto größer ist demnach auch die gemessene ökonomische Ungleichheit. Der Gini-Koeffizient für Welle I lag dabei bei 0,76, der für Welle II bei 0,73, die Vermögen in Österreich sind demnach in der zweiten Welle etwas gleichmäßiger verteilt als in der ersten Welle des HFCS. Aufgrund des Vergleiches der Struktur der Daten aus Welle I und II, speziell der fehlenden Beobachtungen an der Spitze der Daten, ist diese Aussage aber als sehr unzuverlässig einzustufen.

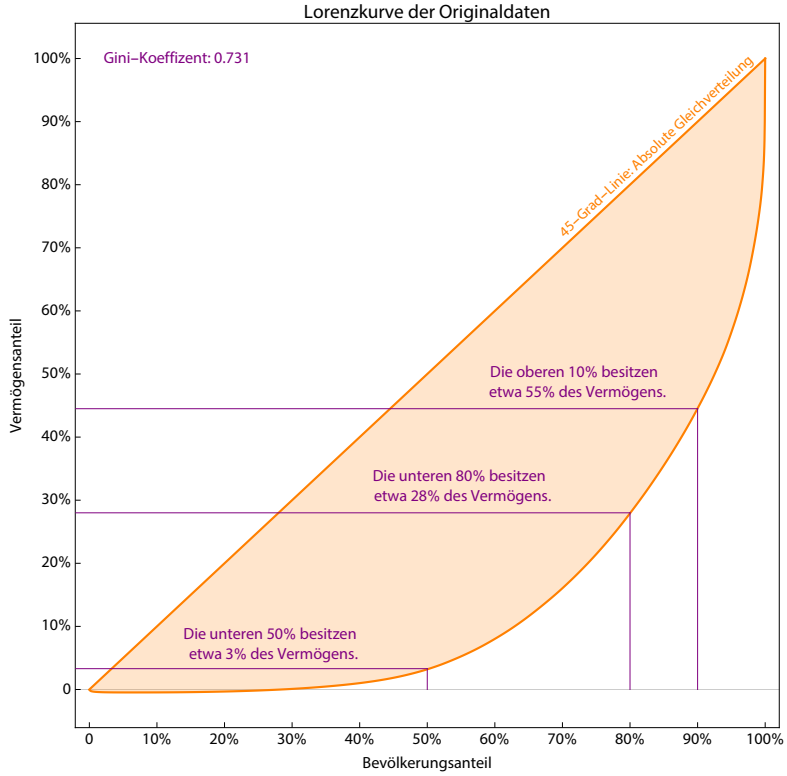


Abbildung 4: Lorenzkurve auf Basis der HFCS-Daten (Originaldaten HFCS II)

3.2. Die Erfassung vermögender Haushalte im HFCS: Ergebnisse aus beiden Wellen

Die in Kapitel 3.1 dargestellten Veränderungen weisen insgesamt auf einen leichten Rückgang der gemessenen Ungleichheit hin, da sich sowohl der Unterschied zwischen Median und Mittelwert als auch der Gini-Koeffizient in der zweiten Welle verringern. Ein genauerer Blick zeigt, dass die Daten der ersten Welle im obersten Dezil signifikant höhere Vermögenswerte aufweisen als die Daten der zweiten Welle, während unterhalb des reichsten Dezils der umgekehrte Fall gilt (vgl. Abbildung 5).

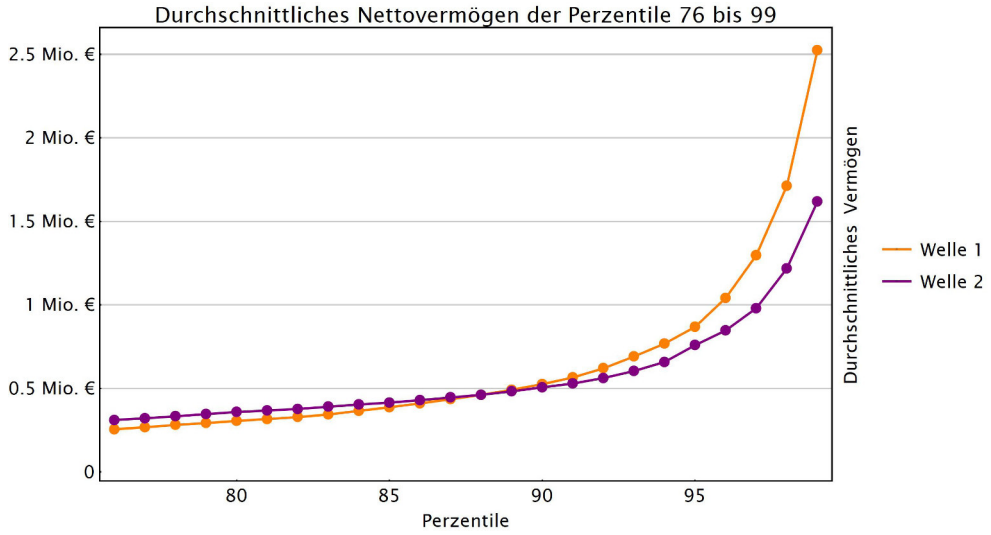


Abbildung 5: Durchschnittliches Nettovermögen in den Perzentilen 76-99 – Österreich Welle I und II

Die gleichmäßigere Verteilung der Vermögen ist demnach durch zwei Faktoren zu erklären: zum Einen wachsen die Vermögen der unteren 90% leicht an, zum Anderen gehen die Vermögen der reichsten 10% der Haushalte stark zurück (vgl. Tabelle 3).

Tabelle 3: Ausgewählte deskriptive Statistiken zu beiden HFCS Wellen für Österreich

	Welle I	Welle II	Differenz
Vermögen Top 10%	617,2 Mrd.	554,9 Mrd.	-62.3 Mrd.
Vermögen Bottom 90%	383 Mrd.	443,3 Mrd.	60.3 Mrd.
Vermögen Bottom 50%	27 Mrd.	32 Mrd.	5 Mrd.
Anzahl Millionäre	174.552	129.304	-45.248
Vermögen Avg.	265.034	258.414	-6,620
Vermögen Total	1.000,2 Mrd.	998,1 Mrd.	-2.1 Mrd.

Während das Wachstum des Vermögens der unteren 90% vor dem Hintergrund der allgemeinen wirtschaftlichen Entwicklung sowie der Entwicklung der Immobilienpreise nicht völlig überraschend kommt, ist der Rückgang des Vermögens des obersten Dezils sowohl an sich als auch hinsichtlich seiner quantitativen Intensität durchaus überraschend: letzteres schrumpft um 62 Mrd. Euro, während ersteres um rund 60 Mrd. Euro wächst.

Besonders der Rückgang der Zahl der gemessen Millionärshaushalte - die Anzahl der in der Gesamtstichprobe abgebildeten Millionärshaushalte schrumpft von Welle I auf Welle II um knapp ein Viertel oder etwa *45.000 Haushalte* (vgl. Tabelle 3) - scheint nur schwer mit den Ergebnissen der ersten Welle in Einklang zu bringen zu sein. Schließlich ist eine derart massive Abschmelzung von (Top-) Vermögen vor dem Hintergrund der allgemeinen wirtschaftlichen Entwicklung unplausibel, vor allem auch da bereits in der ersten Welle der Anteil der reichsten Haushalte unterschätzt wurde (Eckerstorfer et al. 2016, Vermeulen 2016). Darüber hinaus liegen die Ertragsraten größerer Vermögen tendenziell über den durchschnittlichen Ertragsraten (Piketty 2014: 448) – die asymmetrischen Veränderungen in den gemessenen Vermögensbeständen sind damit allerdings kaum in Einklang zu bringen.

Die OeNB verweist in diesem Zusammenhang (OeNB 2016a: 4-5) auf die steigende Bedeutung von *non-response* und *underreporting* – ein Erklärungsansatz, der vor allem vor dem Hintergrund der

angestiegenen Verweigerungsrate (siehe Tabelle 1) im Zuge des Surveys naheliegt. Dabei ist davon auszugehen ist, dass die Wahrscheinlichkeit zur Antwortverweigerung bei vermögenderen Haushalten zunimmt (D'Alessio and Faiella 2002, Kennickell 2008, Kennickell and McManus 1993, OeNB 2016a, Osier 2016, Singer 2006) – ein Zusammenhang, der sich im Zuge der zweiten Welle des HFCS offensichtlich verstärkt hat, wie auch der Abfall der Effective Oversampling Rate nahelegt. Die dokumentierte höhere Verweigerungsrate in Kombination mit einem Anstieg der Vermögenssensitivität bei non-response ist daher geeignet den Rückgang des gemessenen Vermögens an der Spitze der Verteilung zu erklären.

Auf Grund dieses Befunds ließe sich der Rückgang des Gesamtvermögens in der zweiten Welle vor allem durch verändertes Teilnahme- und Antwortverhalten im oberen Teil der Verteilung erklären. Damit ergeben sich allerdings neue Herausforderungen für die Entwicklung adäquater statistischer Schätzverfahren – so fokussieren etwa Eckerstorfer et al. (2013, 2016) vor allem auf das Problem des *non-observation* bias und lassen den Einfluss eines möglichen *non-response* bias außer Acht. Während die Annahmen von Eckerstorfer et al. aufgrund der geringeren Verweigerungsrate in der ersten Welle näherungsweise plausibel scheinen, ist ex-ante unklar inwieweit eine reine Replikation ihrer Anwendung vor dem Hintergrund gestiegener Antwortverweigerungsraten überhaupt sinnvoll ist. Aus diesem Grund nimmt das nachfolgende Kapitel einen Vergleich unterschiedlicher Schätzverfahren mittels Monte Carlo Simulationen vor um jene Verfahren zu bestimmen, die unter gegebenen Umständen die präzisesten Ergebnisse liefern.

4. Zur statistischen Eignung unterschiedlicher Schätzverfahren

4.1. Schätzvarianten

Auf Basis der in Kapitel 2 und 3 aufgeworfenen Problematik einer Verschlechterung der Datenqualität im Rahmen der zweiten Welle des HFCS, stellt sich nun die Frage nach den technischen Möglichkeiten zur Schätzung der entsprechenden Verteilungsfunktion. Dabei orientieren wir uns im Folgendem an dem von Eckerstorfer et al. (2013, 2016) entwickelten Verfahren, das im Wesentlichen darauf beruht (1) die Parameter der Pareto-Verteilung anhand der HFCS-Daten zu bestimmen, (2) die Beobachtungsdaten über einem gewissen Schwellenwert (in Eckerstorfer et al. 2013, 2016: 4 Mio. €) aus dem Datensatz zu entfernen und (3) durch aus der Pareto-Verteilung gewonnene Werte zu ersetzen. Mit Hilfe des so gewonnenen modifizierten Datensatz lassen sich alle relevanten Schätzwerte zum Vermögen in Österreich von neuem berechnen.

Diesem Ansatz folgend betrifft die Frage nach der adäquaten statistischen Bestimmung der Verteilungsfunktion vor allem den ersten der oben angeführten Schritte. Dabei finden sich in der Literatur unterschiedliche Möglichkeiten zur Schätzung einer solchen Verteilungsfunktion, die sich in drei wesentlichen Punkten unterscheiden:

- Die Art des verwendeten Schätzverfahrens: Hier kommen typischerweise Quasi-Maximum-Likelihood Schätzer (ML) und OLS basierte QQ-Schätzungen (Kratz und Resnick 1996) zum Einsatz.
- Die Verwendung von Sampling-Gewichten: Diese werden in der Praxis oft ignoriert, müssen bei Datensätzen mit starkem oversampling bzw. der Verwendung von Reichenlisten jedoch berücksichtigt werden.
- Die Ergänzung der Daten um weitere Informationsquellen, insbesondere diversen Medien entnommenen Reichenlisten: Derartige Listen besonders vermögender Haushalte werden vor allem dann verwendet, wenn die Beobachtungsdaten am oberen Rand von starken Verzerrungen betroffen sind.

Grundsätzlich ist zu beachten, dass die Beobachtungen des HFCS Datensatzes mittels ihrer Gewichtung eine bestimmte Anzahl an Haushalten in der Zielpopulation repräsentieren. Werden nun Beobachtungen aus Reichenlisten für eine Schätzung verwendet, so muss bedacht werden, dass diese Beobachtungen jeweils bloß einen Haushalt repräsentieren und sich in dieser Eigenschaft fundamental von den Beobachtungen des HFCS unterscheiden. Sobald Reichenlisten in der Schätzung verwendet werden, ist es demnach nötig, die Sampling-Gewichte zu berücksichtigen. Während dieser Aspekt in vorhergehenden Arbeiten zur ersten Welle des österreichischen HFCS (aufgrund des Verzichts auf die Verwendung von Reichenlisten) relativ problemlos ignoriert werden konnte (Eckerstorfer et al. 2013, 2016), ist dies im Fall der zweiten Welle nicht möglich, da es mit Hinblick auf das obig dokumentierte veränderte Antwortverhalten nicht zielführend erscheint die Verwendung von Reichstenlisten ex-ante auszuschließen. Die nachfolgenden MonteCarlo-Simulation vergleichen daher im Wesentlichen vier Szenarien:

1.) QQ-Schätzer, Berücksichtigung von Sampling-Gewichten, keine Berücksichtigung einer Reichenliste
2.) QQ-Schätzer, Berücksichtigung von Sampling-Gewichten, Berücksichtigung einer Reichenliste
3.) Quasi-Maximum-Likelihood Schätzer, Berücksichtigung von Sampling-Gewichten, keine Berücksichtigung einer Reichenliste

4.) Quasi-Maximum-Likelihood Schätzer, Berücksichtigung von Sampling-Gewichten, Berücksichtigung einer Reichenliste

4.2. Monte-Carlo Simulationen

Um die relative Eignung der unterschiedlichen Schätzmethoden besser beurteilen zu können und das geeignetste Schätzverfahren auszuwählen, wurden Monte-Carlo Simulationen durchgeführt. Bei einer Monte-Carlo Simulation handelt es sich grundsätzlich um in großer Zahl wiederholte, simulierte Zufallsexperimente, die dazu dienen, analytische Probleme der Wahrscheinlichkeitstheorie numerisch zu lösen. Dies lässt sich in folgendem Versuchsaufbau veranschaulichen: In einem ersten Schritt wird eine synthetische Vermögensverteilung aus N = 200.000 Haushalten generiert, deren Vermögen bei 100.000 Euro beginnt und ab diesem Punkt einer Pareto-Verteilung von α = 1,3 folgt. In einem zweiten Schritt werden insgesamt 10.000 Befragungen dieser Population simuliert, wobei die Größe der Stichprobe zwischen 0,1‰ und 5‰ der generierten Population variiert und die Möglichkeit der Antwortverweigerung gegeben ist. Pro Stichprobengröße werden dabei 200 simulierte Befragungen durchgeführt – da die Stichprobengröße jeweils in Schritten von 0,1‰ variiert wird und somit fünfzig unterschiedliche Stichprobengrößen untersucht werden ergibt sich eine Gesamtzahl von 10.000 Befragungen (200*50). Die Frage ist nun, wie gut unterschiedliche Schätzvarianten den wahren Wert dieser simulierten Vermögensverteilung unter gegebenen Annahmen abbilden können.

Zum Vergleich, die Nettostichprobe des HFCS II für Österreich umfasst 2.997 Haushalte. Bei einer Population von 3.862.526 lt. Gewichten entspricht dies 0.78‰ der Bevölkerung, liegt also im unteren Fünftel der Stichprobengrößen, wie sie in den Monte-Carlo Simulationen untersucht werden.

Auf Basis dieser simulierten Datensätze werden die Gesamtvermögen der jeweiligen Population geschätzt – die Ergebnisse werden in Folge für jede Stichprobengröße in zehn Dezile (je 20 Befragungen pro Dezil) gereiht und geplottet um die Streuung möglicher Ergebnisse zu veranschaulichen. Dabei beschreibt die horizontale Achse die jeweilige Stichprobengröße und die vertikale Achse den geschätzten Vermögenswert (siehe Abbildung 6).

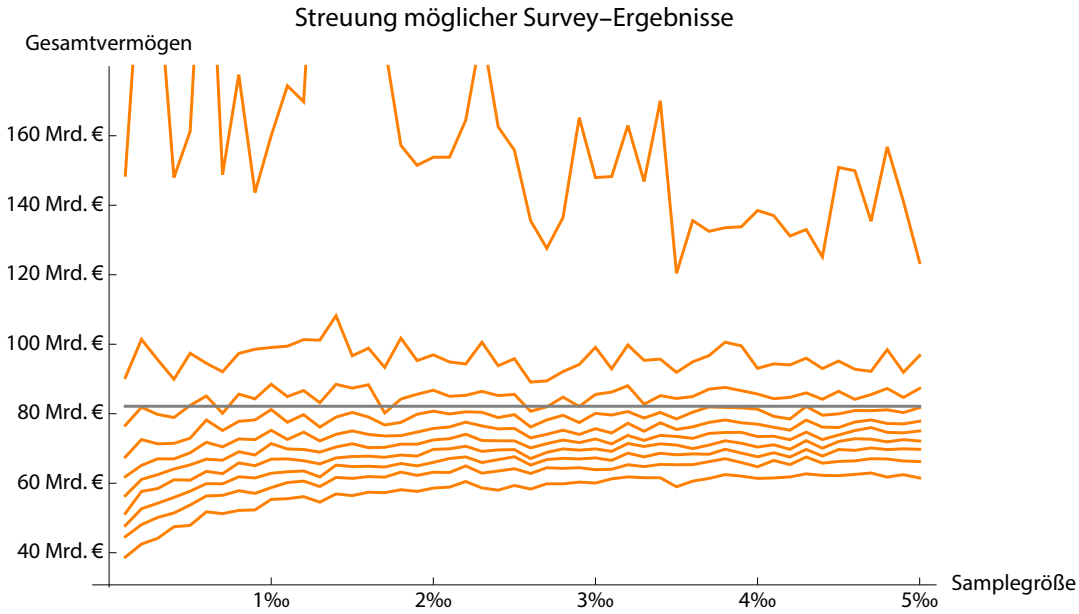


Abbildung 6 :Durchschnitte nach Dezilen der geschätzten Vermögenswerte der Population in Abhängigkeit von der Stichprobengröße

Die unterste Linie in Abbildung 6 zeigt das erste Dezil der durchschnittlich geschätzten Gesamtvermögen für eine zunehmende Größe der Stichproben. Entsprechend stellt die oberste Linie das durchschnittlich geschätzte Gesamtvermögen für das 10te Dezil dar. Bei 200 simulierten Stichprobenziehungen (Befragungen) entspricht somit jede Linie dem Durchschnitt aus 20 Befragungen. Die graue horizontale Linie entspricht dem tatsächlichen Vermögensstand der zu Grunde liegenden Population. Es zeigt sich, dass unabhängig von der Stichprobengröße die unteren 7 Dezile das Gesamtvermögen unterschätzen. Eine Umfrage, die rein auf einer Zufallsstichprobe basiert, unterschätzt somit in 7 von 10 Fällen den wahren Wert. Gleichzeitig ist sichtbar, dass Schätzungen im 10ten Dezil das Gesamtvermögen enorm überschätzen – hier wurden besonders reiche Haushalte in die Stichprobe gezogen, die die entsprechenden Schätzwerte nach oben verzerren. Zu sehen ist darüber hinaus, dass sich die unteren Dezile mit wachsender Stichprobengröße dem wahren Wert leicht annähern. Dies bedeutet grundsätzlich, dass wenn Vermögen auf Basis von Zufallsziehungen geschätzt wird, die Chance hoch ist, dass der tatsächliche Wert des Vermögens unterschätzt wird.

Während Abbildung 6 die Streuung der möglichen Befragungsergebnisse überblicksmäßig darstellt, wird im Folgenden die Performance unterschiedlicher Schätzverfahren analysiert. Konkret untersucht werden die Performance des Quasi-Maximum-Likelihood Schätzers (ML) und des QQ-Schätzers, jeweils mit bzw. ohne Inkorporation einer Reichenliste. Die Performance dieser vier Varianten wird dabei mit jenen Ergebnissen verglichen, die sich aus den unangepassten Befragungs-Daten ergeben. In den untenstehenden Abbildungen werden die Schätzergebnisse des ML (orange) sowie des QQ Schätzers (violett) verglichen und dabei auch die aus den Erhebungsdaten bestimmten Werte (schwarz) sowie die „wahren“ Wert des synthetischen Datensets (graue Linie) angeführt. Für jede Variante sind das 20te und 80te Perzentil der erzielten Ergebnisse sowie der Median angegeben um die Bandbreite der möglichen Resultate in Abhängigkeit von den Eigenschaften der tatsächlichen Stichprobe abzubilden.

Abbildung 7 basiert auf einer Simulation ohne non-response und ohne Reichenliste bei einer Population von 200.000. In diesem Szenario, das im Wesentlichen den Annahmen von Eckerstorfer et al. (2013, 2016) entspricht, ist die Verwendung eines Maximum-Likelihood-Schätzers zu empfehlen, da dieser Ansatz eine viel bessere Annäherung an den wahren Wert ermöglicht, speziell in kleinen Stichproben. Zwar sind grundsätzlich beiden Methoden geeignet, allerdings kann die Verwendung eines QQ-Schätzers unter Umständen zu einer signifikanten Überschätzung des Vermögens führen, weswegen der von Eckerstorfer et al. (2013, 2016) verwendete Maximum-Likelihood Schätzer für diesen Anwendungsfall als geeigneter anzusehen ist.

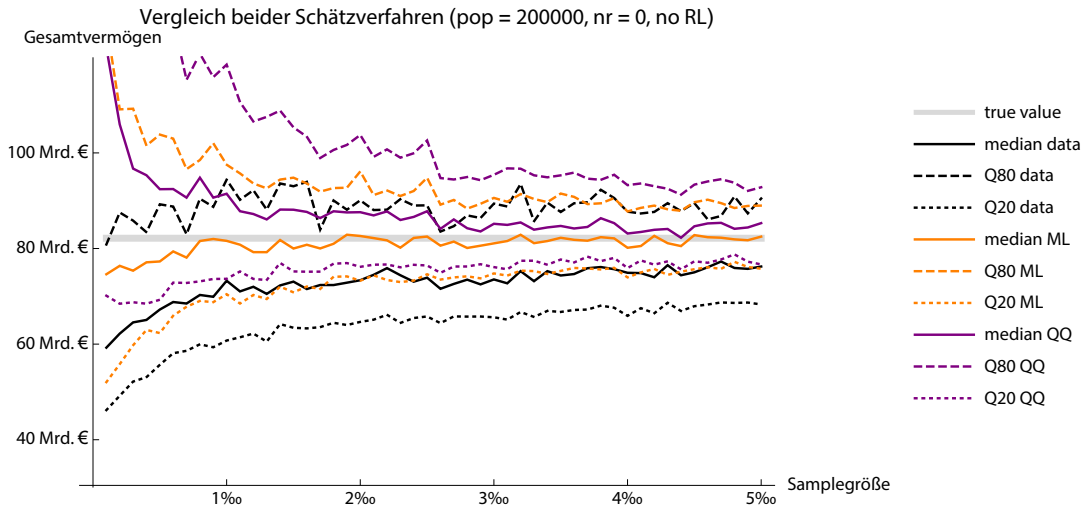


Abbildung 7: Vergleich der Leistungsfähigkeit unterschiedlicher Schätzmethoden anhand von simulierten Stichprobenziehungen ohne non-response und ohne Integration einer Reichenliste

In der nachfolgenden Abbildung 8 gilt grundsätzlich derselbe Versuchsaufbau wie zuvor – mit der Ausnahme, dass hier eine (geringe) non-response-Neigung von 0,2 für den vermögendsten Haushalt vorliegt, die linear zurückgeht und beim ärmsten[7] Haushalt schließlich bei 0 liegt. Es wird also von einer linear fallenden non-response Wahrscheinlichkeit ausgegangen, die für den reichsten Haushalt in der Stichprobe 20% beträgt. Dies entspricht einem sehr konservativem Wert. Zum Vergleich beobachtet der in den USA regelmäßig durchgeführten Survey of Consumer Finances (SCF) für die Welle 2007 eine non-response rate von 65,3% bei der Gruppe der reichen Haushalte, während die restlichen Haushalte zu 32,2% nicht teilnahmen (Kennickell 2010: 6).[8]

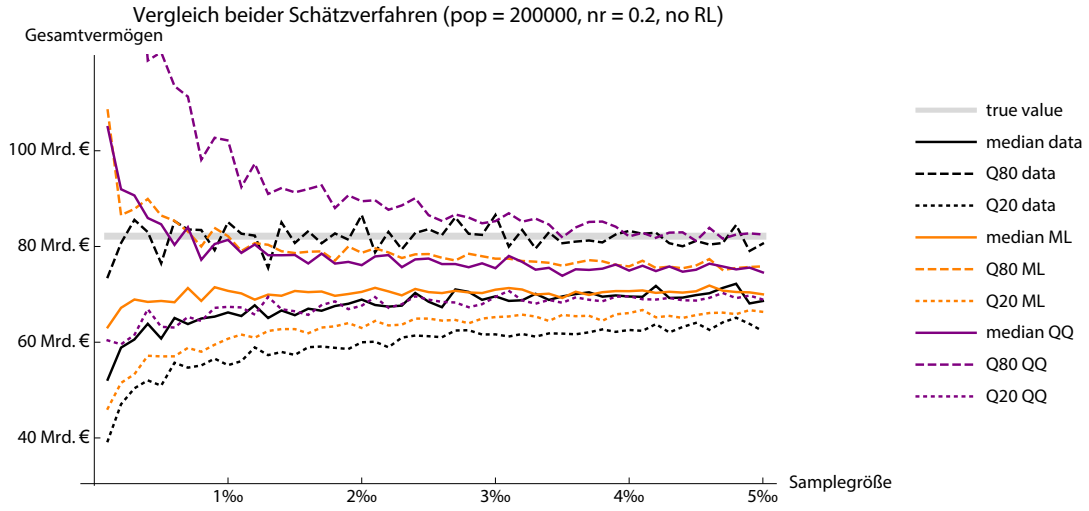


Abbildung 8: Vergleich der Leistungsfähigkeit unterschiedlicher Schätzmethoden anhand von simulierten Stichprobenziehungen mit geringer non-response (0.2) ohne Verwendung einer Reichenliste

In Abbildung 8 ist zu sehen, dass das Schätzergebnis des Maximum-Likelihood Schätzers in etwa jenem auf Basis der Originaldaten entspricht. Der QQ-Schätzer liefert hier Ergebnisse die näher am tatsächlichen Gesamtvermögenswert sind, diesen aber dennoch unterschätzen. Sprich, im Falle von signifikanter non-response produziert der Maximum-Likelihood Schätzer tendenziell problematische Ergebnisse, die weit abseits der korrekten Werte (graue Linie) liegen. Ergo, ist im Fall von nennenswerten Antwortverweigerungsraten, wie sie bei den Daten der zweiten Welle des HFCS für Österreich vorliegen, der QQ-Schätzer zu bevorzugen und liefert grundsätzlich vertrauenswürdigere Ergebnisse. Dabei ist zu erwähnen, dass dieses Ergebnis grundsätzlich robust ist, da auch bei einer Erhöhung der Population bzw. der non-response (0,2-0,8) der QQ-Schätzer bessere Resultate im Median liefert. Robustheitstests zeigen allerdings, dass die Qualität der Ergebnisse beider Schätzmethoden mit steigender non-response rate weiter zurückgeht. Offen ist nun noch die Frage, wie die Qualität der Ergebnisse durch die Ergänzung von Reichenlisten beeinflusst wird.

Abbildung 9 zeigt, dass bei Hinzufügen einer Reichenliste die Ergebnisse des QQ-Schätzers die wahren Werte sehr gut approximieren und vor allem dem Maximum-Likelihood Schätzer sowie einer Schätzung auf Basis der Originaldaten klar überlegen sind.

[7] Anzumerken ist hier, dass der „ärmste“ Haushalt der dargestellten Reihe am Beginn der Pareto-Verteilung liegt – es handelt also nicht um den „ärmsten“ Haushalt der betreffenden Gesellschaft, da die Pareto-Verteilung nur die den oberen Rand der gesamten Verteilung beschreibt (siehe Kapitel 2).

[8] Osier (2016:13) bestimmt für einige teilnehmende Länder des HFCS die response-rates von Haushalten, bei der die Einschätzung des Interviewers berücksichtigt wurde, ob es sich hierbei um bessergestellt oder ärmere Behausungen handelt. Diese hat ergeben, dass in den meisten Ländern die Klassifikation als „bessere gestellte Behausung“ und „bessergestellte“ Nachbarschaft mit einer generellen Abnahme der response-rates einhergeht. In demselben Bericht genannte Informationen der Nationalbank Spaniens zeigt eine abnehmende Kooperationsrate spanischer Haushalte mit steigender Vermögensklasse.

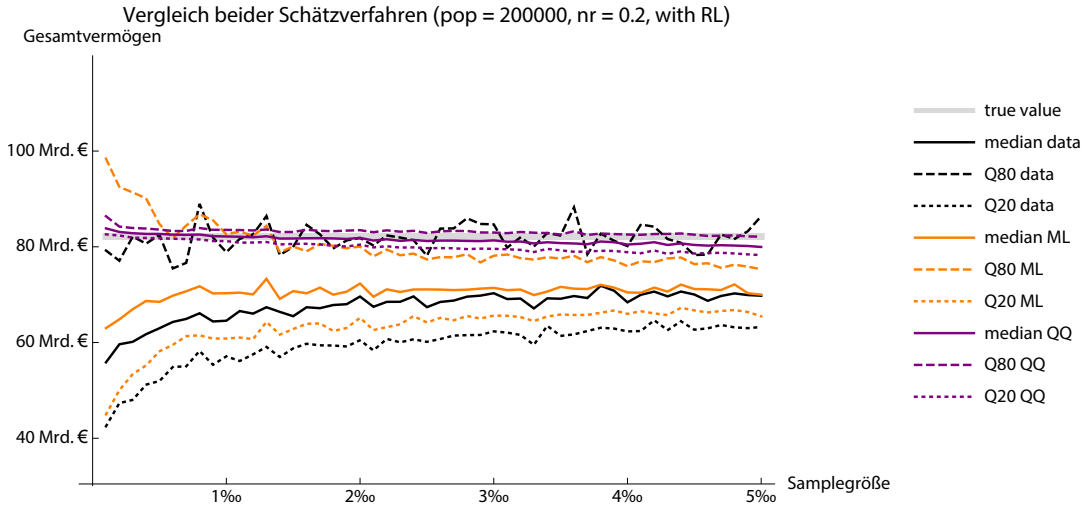


Abbildung 9: Vergleich der Leistungsfähigkeit unterschiedlicher Schätzmethoden anhand von simulierten Stichprobenziehungen mit geringer non-response (0.2) und mit Integration einer Reichenliste

Eine Erhöhung der non-response rate auf 0,8 ist in Abbildung 10 dargestellt. Auch hier bleiben die Ergebnisse aus Abbildung 8 stabil: Unter der Bedingung ungleich verteilter Antwortverweigerung ergeben sich bei Verwendung des QQ-Schätzers wesentlich robustere Ergebnisse.

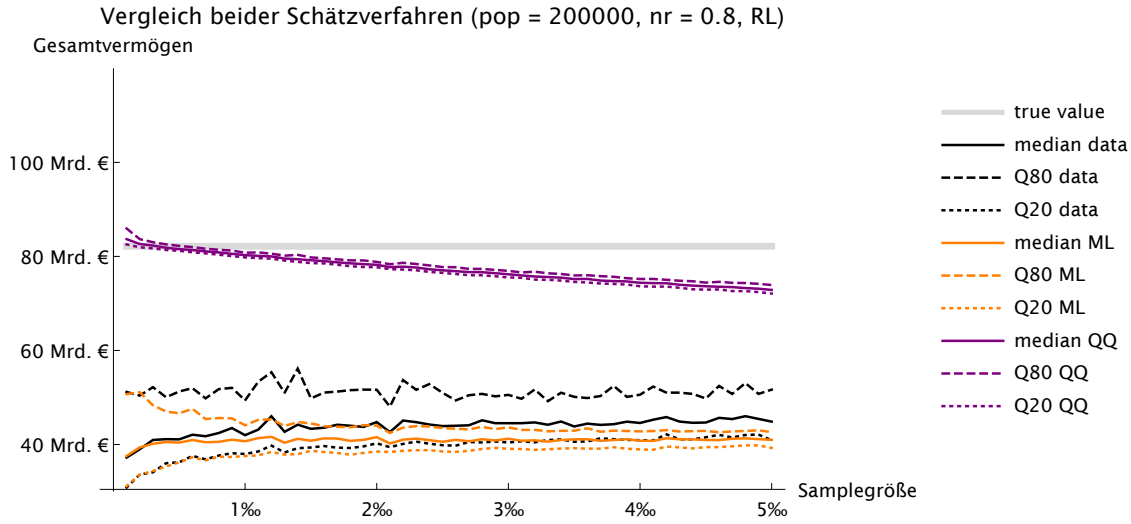


Abbildung 10: Vergleich der Leistungsfähigkeit unterschiedlicher Schätzmethoden anhand von simulierten Stichproben mit starker non-response (0.8) und mit Verwendung einer Reichenliste bei einer Population von 200,000

Zusammenfassend bedeuten diese Ergebnisse, dass es notwendig ist die Schätzmethode anzupassen, da sich das Antwortverhalten bei beiden Wellen der HFCS unterscheidet. Wie oben dargestellt, lässt sich ein Maximum-Likelihood-Schätzer als optimal ausweisen, solange keine nennenswerte, nach Vermögen differenzierte non-response auftritt. Während Eckerstorfer et al. (2013, 2016) mit der Annahme der Absenz signifikanter non-response-Effekte plausible Resultate erzielen konnten, scheint es vor dem Hintergrund der weiter gesunkenen Datenqualität wenig zielführend zu sein diese Annahme weiter aufrecht zu erhalten. Dies bedeutet wiederum, dass für die HFCS Daten der Welle II für Österreich der QQ-Schätzer in Kombination mit einer Reichenliste die vielversprechendsten Ergebnisse liefert. Eine Anpassung der Daten der ersten Welle mit Hilfe der hier diskutierten Methode (i.e. QQ-Schätzer

unter Einbeziehung einer Reichenliste), liefert sehr ähnliche Ergebnisse wie jene in Eckerstorfer et al. (2013, 2016). Dies verdeutlicht einerseits die Robustheit vergangener Schätzungen, verdeutlicht aber andererseits die Notwendigkeit die Analysemethode für Welle II anzupassen. Die einzelnen Arbeitsschritte und Ergebnisse dieser angepassten Methode werden im anschließenden Kapitel dargestellt.

5. Schätzung von privaten Vermögenswerten

Nachdem die Ergebnisse des Verfahrens von Eckerstorfer et al. (2013, 2016) grundsätzlich validiert worden sind und das darin verwendete Schätzverfahren hinsichtlich der Berücksichtigung nicht-gleichverteilter non-response Effekte angepasst wurde, wird dieses im Folgenden auf die Daten der zweiten Welle des HFCS angewendet. Hierzu werden die einzelnen Arbeitsschritte nochmals im Detail beleuchtet. Die notwendigen Schritte der Programmierung finden sich im Anhang in der Form eines *Mathematica*-files.

5.1. Erfassung der Vermögensverteilung

Wie in Kapitel 4 gezeigt wurde, ist bei Verwendung des QQ-Schätzers die Hinzuziehung einer Reichenliste anzuraten, da diese die Qualität und Präzision der Ergebnisse nochmals verbessert. Daher werden in einem ersten Schritt Daten der Trend-Reichenliste (Bezugsjahr 2014; parallel zur Durchführung der Erhebung) zu den HFCS-Daten hinzugefügt und mit einem Gewicht von 1 versehen (Gewichte drücken im Rahmen des HFCS die Zahl der repräsentierten Haushalte aus). Da die Durchführung der Schätzung unter Hinzuziehung der Gewichte vorgenommen wird, bleibt der spezifische Charakter dieser Datenpunkte auch formal erhalten. An dieser Stelle ist zu erwähnen, dass in der Trend-Reichenliste Familienclans angeführt werden, die vor der Verwendung in einzelne Haushalte gespalten werden. Es ergeben sich dabei sieben zusätzliche Haushalte.[9]

Im Gegensatz zu Arbeiten die den Schwellenwert der Pareto-Verteilung von Vermögen schlicht annehmen (vgl. Bach et al. 2012, Bach und Beznoska 2012), wird dieser im Zuge der Methode von Eckerstorfer et al. (2013, 2016) auf Basis statistischer Kriterien bestimmt. Dabei werden Vermögenswerte an den ersten 30 Perzentilgrenzen der Vermögensverteilung als mögliche Schwellenwerte für die Pareto-Verteilung in Betracht gezogen. In einem nächsten Schritt werden mit Hilfe des QQ-Schätzers an all diesen Punkten der Formparameter α der Pareto-Verteilung bestimmt (vgl. Clauset et al. 2009). Es ergeben sich so über alle fünf Imputationen der Daten für die obersten 30 Perzentile geschätzte Paretoverteilungen. Diese geschätzten Verteilungen werden in Folge mittels eines Cramer-Von-Mises Tests dahingehend geprüft, inwieweit sie mit der Verteilung der vorliegenden Daten übereinstimmen. Dieser beruht im Wesentlichen auf der Idee, den Erwartungswert des quadrierten Unterschieds zwischen der geschätzten hypothetischen Verteilung und der empirischen Verteilungsfunktion der Daten zu berechnen. Anhand dieses Tests wird jenes Schätzergebnis bestimmt, dass die geringste Abweichung zu den gegebenen Daten aufweist. Da die Null-Hypothese des Cramer-von-Mises Tests lautet, dass die Daten der Verteilung entsprechen, sind jene Punkte zu wählen, die die niedrigsten Teststatistiken ausweisen. Die folgenden zwei Grafiken zeigen zum einen die Werte der geschätzten Pareto-Aphas (Steigungsparameter) für alle fünf Imputationen der Daten über die obersten 30 Perzentile, zum anderen die für selbige anhand des Cramer-von-Mises Tests errechneten Teststatistiken.

[9] Siehe dazu auch den Online Appendix von Eckerstorfer et al. (2016).

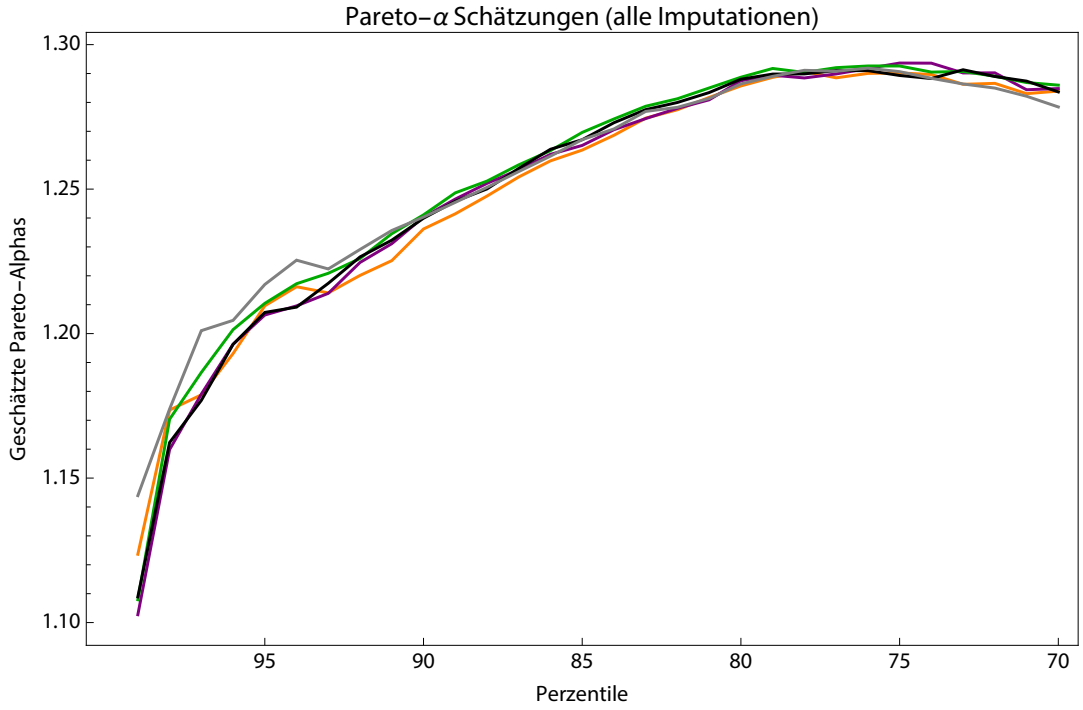


Abbildung 11: Pareto-Alpha Parameter über fünf Imputationen und 30 Perzentile

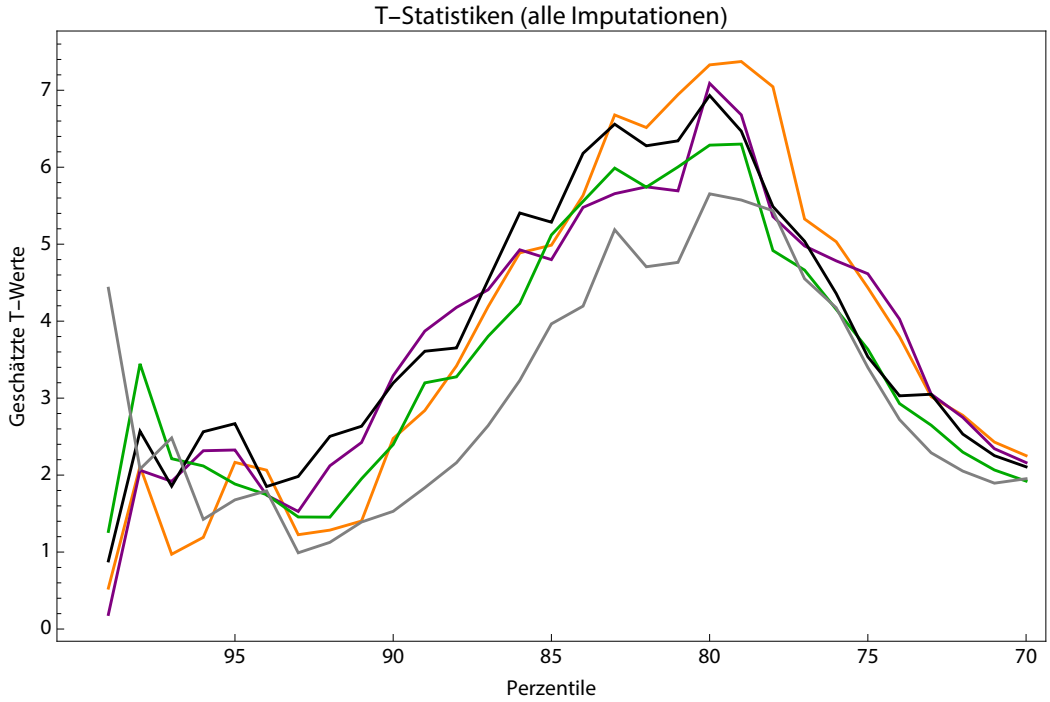


Abbildung 12: Cramer von Mises Teststatistiken über fünf Imputationen und 30 Perzentile

Die daraus resultierenden Teststatistiken werden mit Hilfe einer Maxi-Min Analyse (Wald 1945) untersucht um das geeignetste Perzentil zu ermitteln, ab dem von einer Paretoverteilung in den Daten gesprochen werden kann. Die Datenpunkte oberhalb dieses Ansatzpunktes bilden dann die Basis für die Schätzung der Paretoverteilung. Als Schätzungsparameter ergeben sich auf Basis dieser Vorgangsweise das 93. Perzentil, sowie ein Pareto-Alpha von 1,218. Zum Vergleich, in Eckerstorfer et al. ergaben sich als respektive Werte das 78. Perzentil und 1,277. Die nachstehende Tabelle zeigt die geschätzten Werte für die Pareto-Alphas sowie die Nettovermögen der jeweiligen Ansatzpunkte (Schwellenwert des 93. Perzentils) über alle fünf Imputationen.

Tabelle 4: Nettovermögen und geschätzte Pareto-Alphas am Schwellenwert des 78. Perzentils

Imputation	Pareto-Alpha	Ansatzpunkt für Paretoverteilung
1	1,21408	Nettovermögen von € 620.700
2	1,21393	Nettovermögen von € 617.480
3	1,22088	Nettovermögen von € 630.800
4	1,2174	Nettovermögen von € 617.515
5	1,22237	Nettovermögen von € 642.700
Durchschnitt	1,21773	Nettovermögen von € 625.839

5.2. Datenanpassung

Die Schätzung des nicht erhobenen oberen Verteilungsrands auf Basis der oben errechneten Parameter erfolgt in folgenden Schritten. Zuerst werden alle Beobachtungen mit einem Nettovermögen von > 4.000.000 Euro aus dem Datensatz entfernt. Dies entspricht durchschnittlich acht Beobachtungen pro Imputation. Im nächsten Schritt wird die zuvor geschätzte Pareto-Verteilung herangezogen um die Anzahl der Haushalte (H_i) zu errechnen, die ein Vermögen > 4.000.000 Euro aufweisen. Hierzu wird aus dem HFCS-Datensatz die Anzahl der Haushalte (HH_i) berechnet, welche zwischen dem jeweiligen Ansatzpunkt (m_i) für die Pareto-Verteilung und der 4-Millionen-Grenze (μ) liegen. Die Zahl der Haushalte mit einem Nettovermögen größer als 4 Millionen Euro (H_i) kann mithilfe der Pareto--Verteilung schließlich wie folgt berechnet werden:

$$H_i = HH_i \frac{1-P_i(\mu)}{P_i(\mu)} \qquad (2)$$

Im nächsten Schritt wird die so berechnete Anzahl an Haushalten mit einem Nettovermögen größer als 4 Millionen Euro mit Hilfe der Pareto-Verteilung generiert, wobei die Anzahl der zu generierenden Haushalte von Imputation zu Imputation unterschiedlich ist. Hierbei kann das Vermögen (x_i) eines jeden Haushaltes oberhalb der 4-Millionen-Grenze wie folgt berechnet werden:[10]

$$x_i = m_i \left(\frac{HH_i+H_i}{H_{x_i}}\right)^{1/\alpha_i} \qquad (3)$$

Die generierten Beobachtungen werden in Folge dem HFCS-Datensatz mit einem Gewicht von 1 hinzugefügt. Jede repräsentiert also exakt einen Haushalt. Zuletzt müssen die - durch die vorgenommene Anpassung nicht mehr 100% kohärenten - Gewichtungen der Original-Haushalte korrigiert werden. Dabei wird die Nettoveränderung innerhalb der jeweiligen Imputation in Relation zur jeweiligen Gesamtbevölkerung gesetzt und zur Korrektur die jeweiligen Gewichte linear abgeschmolzen bzw. aufgewertet. Die generierten Vermögenswerte wurden in diesem Prozess mit einer Milliarde Euro

[10] Dies folgt aus dem Umstand, dass $1 - P_i(x_i) = \Pr(X_i > x_i) = (m_i/x_i)^{\alpha_i} \equiv \frac{H_{x_i}}{HH_i+H_i}$.

gedeckelt. Dies liegt einerseits an einer gewissen Präferenz für konservative Berechnungen, sowie andererseits an einer gewissen Skepsis dahingehend, dass die vorliegenden statistischen Schätzungen eine solide Grundlage für Aussagen über eine so kleine Gruppe von Haushalten am äußersten Rand der Verteilung bieten.

5.3. Verteilungsstatistik auf Basis angepasster Daten

Basierend auf den, von obigen Simulationen gestützten, Schätzparametern (QQ-Schätzer, mit Gewichten, mit Reichenliste) ergeben sich folgende angepassten Werte für den österreichischen Vermögensbestand (in untenstehender Tabelle im Vergleich zur Schätzung ohne eine Paretoannahme basierend auf Daten der HFCS II und den Resultaten von Eckerstorfer et al. (2016) zur HFCS I):

Tabelle 5: Schätzungsergebnisse

Vermögensschätzung	Originaldaten HFCS II	**Pareto-Methode - Daten HFCS II**	Eckerstorfer et al. (2016) HFCS I
Durchschnittsvermögen	258k	**341k**	339k
Gesamtvermögen	998 Mrd.	**1.317 Mrd.**	1.278 Mrd.
Anteil Top 1%	25%	**41%**	38%
Anteil Top 5%	43%	**56%**	59%
Anteil Top 10%	56%	**66%**	69%
Anteil Top 20%	72%	**79%**	82%
Anteil Unterste 50%	3,2%	**2,5%**	2,2%
Anzahl MillionärInnen	129k	**148k**	181k
Anzahl MilliardärInnen	0	**35,8**	30,6
Feldforschung	06/2014-02/2015		09/2010-05/2011

Wie zu sehen ist wächst der Stand des Gesamtvermögens um 319 Mrd. Euro, das Durchschnittsvermögen wächst um 83.000 Euro und der Anteil des obersten Prozents steigt von 25% auf 41%. Diese Ergebnisse sind trotz leicht abweichender Berechnungsmethode und unterschiedlicher Datenlage in einer ähnlichen Größenordnung wie die Schätzung mit den Daten der ersten Welle des HFCS. In weiterer Folge werden nun die grafischen Darstellungen aus Kapitel 3 auf Basis der angepassten Vermögensdaten repliziert. Tabelle 6 zeigt die angepassten Werte für Gesamt- und Durchschnittsvermögen für die obersten fünf Perzentile. Hier wird deutlich, dass der überwiegende Teil des Zugewinns an Vermögen im obersten Perzentil auftritt. Hier wird der Vermögensbestand mehr als verdoppelt. Dahingegen sind die Auswirkungen auf die Ergebnisse für die restliche Bevölkerung eher gering. Ein Vergleich von Abbildung 13 und Abbildung 2 zeigt hier, dass sich an der Verteilung der Haushalte auf die verschiedenen Vermögensklassen nicht viel ändert, da der überwiegende Teil der Veränderung innerhalb der Klasse der Haushalte mit einem Nettovermögen < € 500.000 auftritt.

Die Lorenzkurve in Abbildung 14 zeigt wiederum eine signifikante Verschärfung der ökonomischen Ungleichheit durch die Anpassung der HFCS II Daten.

Tabelle 6:Vermögensverteilung der obersten 5 Perzentile HFCS II - Schätzung

Perzentil	Gesamtnettovermögen im Perzentil	Durchschnittsnettovermögen im Perzentil
96	€ 35,2 Mrd.	€ 904.206
97	€ 41,6 Mrd.	€ 1,07 Mio.
98	€ 53,5 Mrd.	€ 1,39 Mio.
99	€ 76,9 Mrd.	€ 2,01 Mio.
100	€ 534 Mrd.	€ 14,05 Mio.

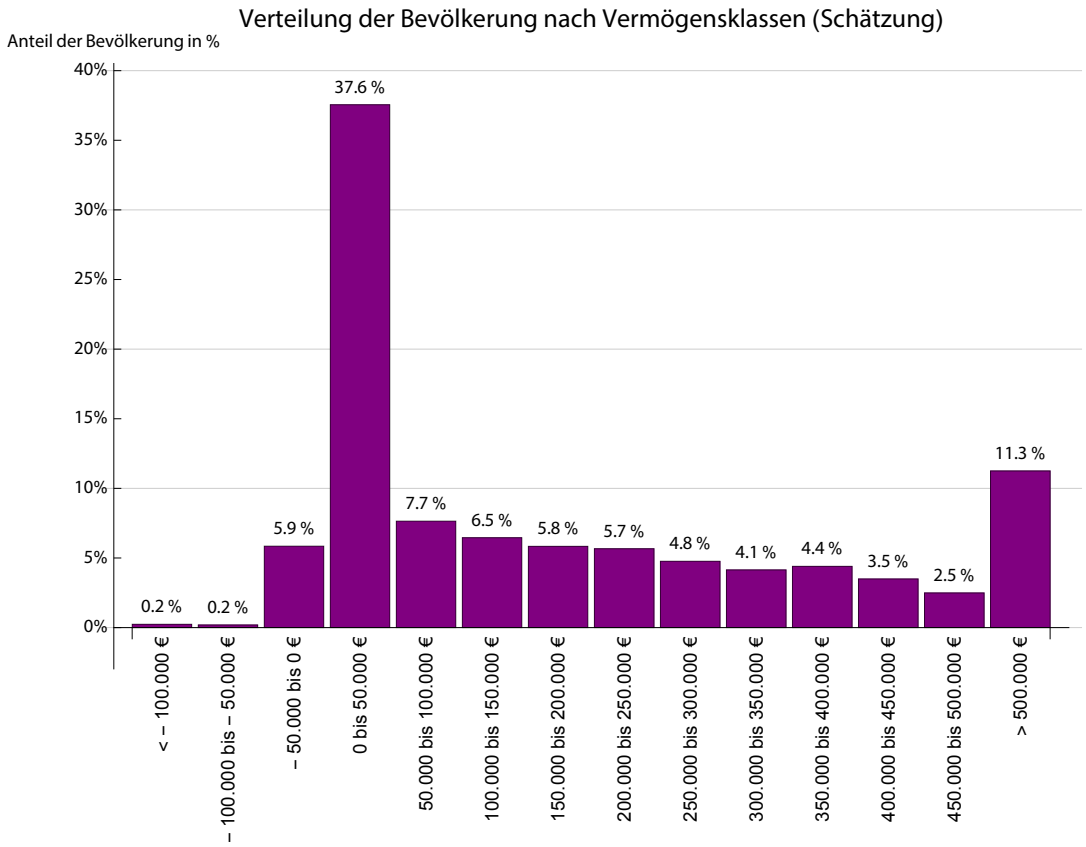


Abbildung 13: Vermögensverteilung in Österreich nach Vermögensklassen (Schätzung)

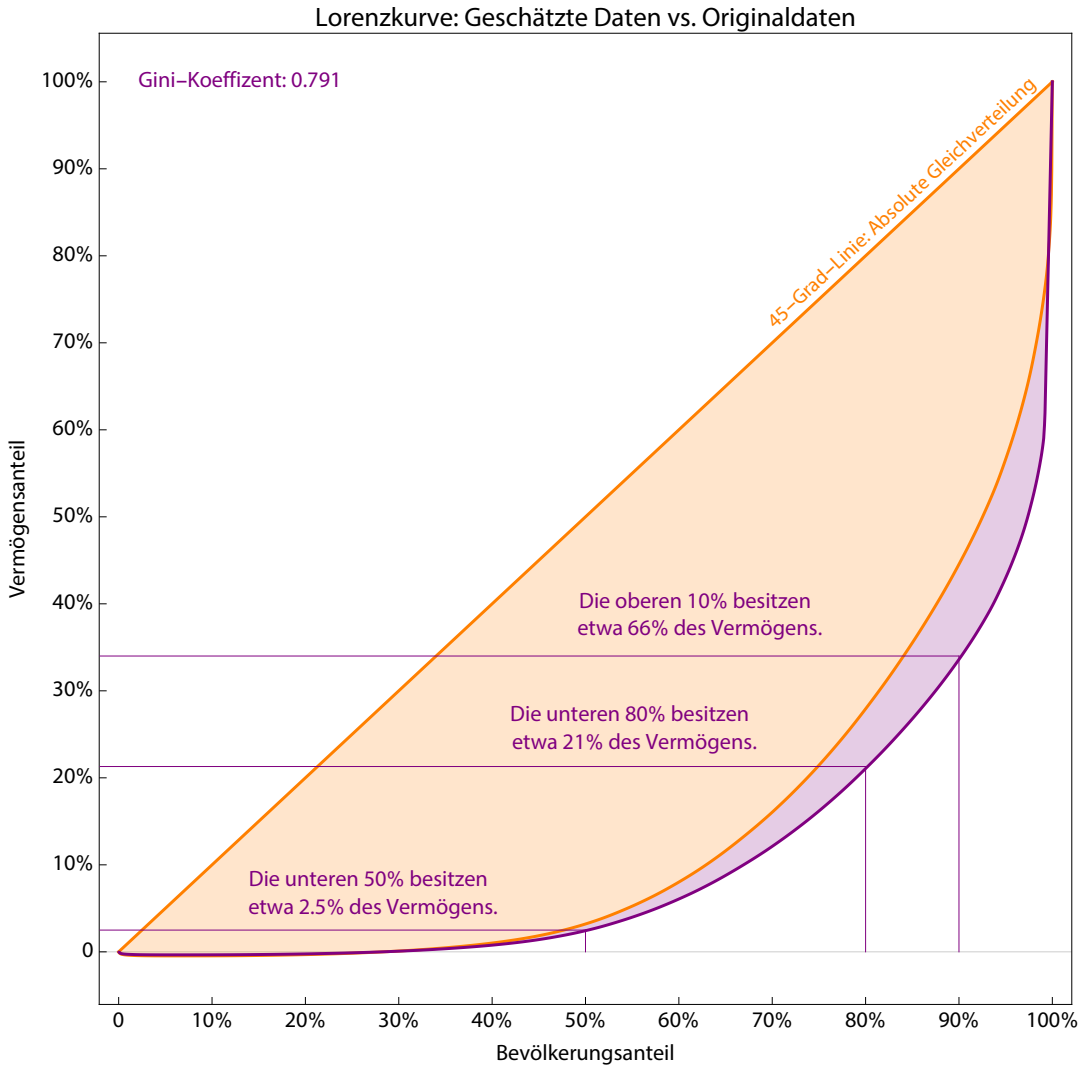


Abbildung 14: Lorenz-Kurve der originalen (orange) und angepassten (lila) HFCS Daten im Vergleich

6. Robustheit der Ergebnisse

Auf Grund der signifikanten Unterschiede zwischen den Implikationen der Originaldaten des HFCS und den angepassten Daten, soll im Folgenden der Grad der Unsicherheit des Schätzverfahrens genauer besprochen werden um die Robustheit der erzielten Resultate besser beurteilen zu können. Hierfür werden zwei allgemein anerkannte Ansätze gewählt: In einem ersten Schritt werden die dem Datensatz beigelegten Replikationsgewichte genutzt um Varianz und Standardabweichung des geschätzten Pareto-Alphas zu bestimmen; dies erlaubt es die Schwankungsbreite bzw. Präzision der originalen Schätzung genauer zu eruieren (Kapitel 6.1). In einem zweiten Schritt werden die zur Schätzung des Pareto-Alpha herangezogenen Daten einem Bootstrap-Verfahren unterzogen um festzustellen, ob das erzielte Ergebnis auch gegenüber Variationen in den zur Verfügung stehenden Daten robust bzw. stabil ist.

6.1. Konfidenzintervalle mit Replicate Weights

Vor dem Hintergrund des Umstands, dass die Daten des HFCS in fünf Imputationen vorliegen, die vornehmlich dazu dienen item non-response Probleme zu beheben, ist es für die Berechnung der Varianz einzelner Schätzer notwendig, die im Zuge der Datenerhebung und des damit verbundenen Stratifizierungsverfahrens auftretenden Unsicherheiten zu berücksichtigen (ECB 2016a: 61). Zu diesem Zweck ist dem HFCS ein entsprechender Datensatz von 1.000 Replikationsgewichten beigelegt. Vor diesem Hintergrund stehen die in diesem Kapitel angestrebten Berechnungen zur Robustheit. Dabei werden die Replikationsgewichte herangezogen um die in der Datenerhebung auftretende Unsicherheiten und damit verbundene Schwankungen in den Ergebnissen zu errechnen. Konkret lassen sich mit Hilfe der genannten Replikationsgewichte 1.000 unterschiedliche Schätzwerte des Alpha-Parameters der Pareto-Verteilung errechnen, die die aus den Unsicherheiten in der Datenerhebung hervorgehenden Schwankungen abbildet. Diese wurden dabei auf zwei Arten ausgewertet: Zum einen wurde die Rangfolge der 1.000 Schätzer verglichen und jener Bereich eingegrenzt, der die mittleren 95% der erhaltenen Schätzwerte repräsentiert (Indikator *Rank* in Tabelle 7). Zum anderen lässt sich mit Hilfe der Replikationsgewichte die Standardabweichung des geschätzten Pareto-Alpha errechnen. Diese kann dazu verwendet werden um ein Intervall von zwei Standardabweichungen rund um den Punktschätzer zu definieren (Indikator *Double SD* in Tabelle 7), wobei hier sowohl die Variation innerhalb der Imputation als auch jene zwischen den Imputationen Berücksichtigung findet (ECB 2016a: 61). In folgender Tabelle sind diese Werte dargestellt; dabei wurden auch die korrespondierenden Schätzungen zum Nettovermögen errechnet.

Tabelle 7: Ergebnisse des Robustheitschecks auf Basis der Replicate Weights

	Lower Bound Double SD	Upper Bound Double SD	Point Estimate	Lower Bound Rank	Upper Bound Rank
Pareto-α	1,318	1,117	1,22	1,304	1,106
Nettogesamtvermögen	1.163 Mrd. Euro	1.562 Mrd. Euro	1.317 Mrd. Euro	1.180 Mrd. Euro	1.597 Mrd. Euro

In Tabelle 7 zeigt sich nicht nur, dass beide gewählten Indikatoren sehr ähnliche Ergebnisse liefern, sondern auch, dass die unteren Schwellenwerte immer noch um fast 200 Mrd. über dem aus den Originaldaten des HFCS errechneten Wert von 998 Mrd. Euro liegen. Das Ergebnis einer

Unterschätzung des Gesamtvermögens in den Originaldaten ist vor diesem Hintergrund als robust einzustufen; darüber weist auch die Symmetrie der errechneten Schwellenwerte auf eine valide Schätzung hin.

Während dieses Kapitel demnach Unsicherheiten auf der Ebene der Datenerhebung an Hand von Replikationsgewichten behandelt hat, beschäftigt sich das nachfolgende Kapitel mit der internen Robustheit der Daten des HFCS II.

6.2. Bootstrap

Eine weitere Methode die oben angegebenen Ergebnisse zu validieren ist die Variabilität der vorhandenen Daten anhand eines „*bootstraps*“ zu analysieren. Generell, wird mittels eines solchen „bootstraps“ überprüft wie sensitiv die Ergebnisse auf Variationen in den verfügbaren Daten reagieren, was einen Robustheitstest für die Ergebnisse darstellt.

Es werden folgende Berechnungen vorgenommen: zuerst werden 1000 zufällige Stichproben aus dem originalen HFCS Datenset gezogen, bestehend aus jeweils 60% der Daten (2997*0,6 =~ 1798) und 45 Haushalten aus der Reichstenliste (von 67). Danach werden die Pareto-Alpha Werte für jede der 1.000 zufällig gezogenen Stichproben berechnet und nach ihrer Größe gereiht. Die aus dem Bootstrap resultierenden oberen und unteren Ränder der möglichen Ergebnisse repräsentieren dabei 95% der gesamten Variation der Ergebnisse. Mit Hilfe dieser Werte lassen sich die nicht erhobenen Topvermögenswerte neuerlich berechnen um das durch diese Pareto-Alpha Parameter vorhergesagt Gesamtvermögen zu bestimmen. Die Ergebnisse dieser Schätzung werden zusammen mit den entsprechenden Alpha Parameter in untenstehender Tabelle 8 angeführt. Das „*Point Estimate*“ entspricht den in Kapitel 5 geschätzten Werten.

Tabelle 8: Ergebnisse eines Robustheitschecks mittels Bootstrap-Verfahren

	Lower Bound	Point Estimate	Upper Bound
Pareto-α	1,24	1,22	1,19
Nettogesamtvermögen	1.279 Mrd. Euro	1.317 Mrd. Euro	1.368 Mrd. Euro

Wie in Tabelle 8 zu sehen ist, ist der „upper bound“ des Gesamtvermögens 50 Milliarden Euro über dem „Point Estimate“, während der „lower bound“ 39 Mrd. Euro weniger Vermögen aufweist. Dieses Ergebnis indiziert, dass die gewählte Schätzmethode robust gegenüber zufälligen Variationen der Stichprobe ist.

7. Aufkommenspotential einer allgemeinen Vermögenssteuer

Abschließend werden die Daten zum Nettovermögen der österreichischen Haushalte aus dem HFCS sowie die mit Hilfe der Pareto-Verteilung angepassten Vermögenswerte (siehe Kapitel 5) dazu verwendet um das Aufkommen einer allgemeinen Vermögenssteuer zu schätzen. Bei diesen Berechnungen werden ausschließlich Haushalte als Steuerobjekte gewertet, jedoch wird das in Körperschaften gebundene Vermögen als Bestandteil des Privatvermögens der Haushalte berücksichtigt. Es werden im Folgenden sechs unterschiedliche Steuertarife auf Basis von vier unterschiedlichen Annahmekombinationen analysiert (vgl. Tabelle 9): Zuerst wird der jeweilige Tarif schlicht auf die unveränderten Originaldaten der HFCS II angewandt. Eine zweite Simulation liefert ein Schätzergebnis für den mit Hilfe der Pareto-Schätzung modifizierten Datensatz. Eine dritte und vierte Anwendung wiederholt die Schätzung auf Basis der modifizierten Daten, berücksichtigt aber Ausweicheffekte seitens der Steuersubjekte. Zur Quantifizierung dieser Effekte wurden in der entsprechenden Literatur etablierte Größen herangezogen (vgl. Bach und Beznoska 2012). Diese drücken aus welcher Anteil der Bemessungsgrundlage sich der Besteuerung entzieht: Immobilienvermögen 20%, Finanzvermögen 24%, Firmenvermögen 13%, andere Vermögenswerte 100%.[11] Schließlich wird in der letzten Variante untersucht, wie sich sehr starkes Ausweichverhalten auf die Aufkommensschätzung auswirkt. Zu diesem Zweck wird bei Finanz- und Firmenvermögen von den doppelten Werten von Bach und Beznoska (2012) ausgegangen (48% bzw. 26%). Dabei stehen im Nachfolgenden vor allem die errechneten Steueraufkommenspotentiale im Vordergrund – es können aber auch jederzeit alternative Indikatoren im Modell bestimmt werden (z.B. Anzahl der betroffenen Haushalte). Schließlich wird in Tabelle 9 auch die statistische Schwankungsbreite der oben besprochenen Schätzer angegeben – hierzu wurde wiederum die Rangfolge der sich bei Verwendung der Replikationsgewichte ergebenden Werte herangezogen. Die Berechnung der Schwankungsbreiten erfolgt also analog zur Bestimmung des Indikators *Rank* in Kapitel 6.1.

Bei den ersten beiden simulierten Tarifen handelt es sich um lineare Steuermodelle mit einer Freigrenze von 500,000 Euro (was in etwa eine Besteuerung des ersten Dezils bedeuten würde) in Modell I und 1 Million Euro in Modell II. Der Steuersatz beträgt jeweils 1%. Bei den anderen vier Modellen handelt es sich jeweils um progressive Tarife. Das erste dieser Modelle besteht aus einem zweistufigen Steuertarif mit einem Freibetrag von 1 Million Euro und eher niedrigen Steuersätzen (0,3% zwischen 1 und 2 Millionen, darüber 0.7%). Die verbleibenden Tarife sind jeweils dreistufig: Variante II hat einen Freibetrag von 700,000 Euro (Steuersätze: 700,000 - 1 Million: 0.5%; 1-2 Millionen: 1%; >3 Millionen: 1,5%) und Variante III von 1 Million Euro (Steuersätze: 1-2 Millionen: 0.7%; 2-3 Millionen: 1%; > 3 Millionen: 1.5%). Variante IV besteht schließlich aus einem Steuermodell, bei dem das Ausmaß der Besteuerung mit steigendem Vermögen überproportional ansteigt. Im Detail sieht dieser Tarif einen Freibetrag bis 2 Millionen Euro vor – ab diesem Punkt würde ein Steuersatz von 1% (bis 10 Millionen), 1.5% (zwischen 10 und 100 Millionen) und 4% (ab 100 Millionen) schlagend werden. Ein letzter Tarif zeigt den Fall eines Freibetrags von 1 Million Euro, einem Steuersatz von 0.5% zwischen 1 und 10 Millionen und 1% für Vermögen > 10 Millionen Euro. Die nachstehende Tabelle 2 zeigt die geschätzten Steueraufkommen unter Berücksichtigung der hier angenommenen Freibeträge und Steuersätze. Die Werte basieren auf dem in den vorgegangenen Kapiteln entwickelten Verfahren.

[11] Hier wird angenommen, dass die relative Bedeutung der einzelnen Vermögenskomponenten im Bereich der mit Hilfe der Pareto-Verteilung neu geschätzten Vermögen am obersten Rand der relativen Bedeutung dieser Komponenten in jenen beobachteten Haushalten entspricht, die im Zuge der Vermögensschätzung aus dem Datensatz entfernt werden (siehe Kapitel 5 für Details).

Tabelle 9: Geschätzte Aufkommen einer allgemeinen Vermögenssteuer

	Originaldaten	Modifizierte Daten	Modifizierte Daten und Ausweicheffekte	Modifizierte Daten und starke Ausweicheffekte
Lineares Modell I *Freibetrag: 500.000 Euro* *Steuersatz: 1%*	3.6 Mrd.	6.7 Mrd. (5.4 - 9.5 Mrd.)	5 Mrd. (3.9 - 7.3 Mrd.)	4.5 Mrd. (3.5 - 6.6 Mrd.)
Lineares Modell II *Freibetrag: 1 Million Euro* *Steuersatz: 1%*	2.5 Mrd.	5.5 Mrd. (4.2 - 8.3 Mrd.)	4.2 Mrd. (3.2 - 6.5 Mrd.)	3.8 Mrd. (2.8 - 5.8 Mrd.)
Progressive Steuer I *Freibetrag: 1 Million Euro* *Steuersatz:* *1-2 Millionen: 0.3%* *> 2 Millionen; 0.7%*	1.5 Mrd.	3.5 Mrd. (2.6 - 5.4 Mrd.)	2.7 Mrd. (2.0 - 4.2 Mrd.)	2.4 Mrd. (1.8 - 3.8 Mrd.)
Progressive Steuer II *Freibetrag: 700.000 Euro* *Steuersatz:* *700.000-2 Mil.: 0.5%* *2 -3 Millionen; 1%* *> 3 Millionen: 1.5%*	3.2 Mrd.	7.5 Mrd. (5.5 - 11.5 Mrd.)	5.7 Mrd. (4.2 - 8.9 Mrd.)	5.1 Mrd. (3.7 - 8.1 Mrd.)
Progressive Steuer III *Freibetrag: 1 Million Euro* *Steuersatz:* *1-2 Millionen: 0.7%* *2-3 Millionen; 1%* *> 3 Millionen: 1.5%*	3 Mrd.	7.4 Mrd. (5.4 - 11.4 Mrd.)	5.7 Mrd. (4.1 - 8.9 Mrd.)	5.1 Mrd. (3.7 - 8.0 Mrd.)
Progressive Steuer IV *Freibetrag: 2 Millionen Euro* *Steuersatz:* *2-10 Millionen: 1%* *10-100 Millionen; 1.5%* *> 100 Millionen: 4%*	2 Mrd.	8.3 Mrd. (5.5 - 14.8 Mrd.)	6.3 Mrd. (4.1 - 11.3 Mrd.)	5.6 Mrd. (3.6 - 10.1 Mrd.)
Progressive Steuer V *Freibetrag: 1 Millionen Euro* *Steuersatz:* *1-10 Millionen: 0.5%* *> 10 Millionen: 1%*	1.7 Mrd.	4.2 Mrd. (3.0 - 6.6 Mrd.)	3.2 Mrd. (2.3 - 5.2 Mrd.)	2.9 Mrd. (2.0 - 4.7 Mrd.)

8. Resümee

Ziel der vorliegenden Arbeit war es, die Untererfassung reicher Haushalte in der zweiten Welle des HFCS unter der Annahme einer Pareto-Verteilung für den oberen Rand der Vermögensverteilung zu korrigieren und damit eine realistischere Darstellung der Bestände und Verteilung privater Vermögen in Österreich zu liefern. Dabei wurden auf Basis von Monte-Carlo Simulationen verschiedene Varianten der Implementierung dieser Pareto-Methode überprüft. Es hat sich gezeigt, dass bei Vorliegen nicht-gleichverteilter Antwortverweigerungen, die insbesondere die Spitze der Vermögensverteilung betreffen, der QQ-Schätzer in Kombination mit einer Liste der reichsten ÖsterreicherInnen gut geeignet ist um den oberen Rand der österreichischen Vermögensverteilung statistisch abzubilden. Die Methode von Eckerstorfer et al. (2013, 2016) wurde damit um Überlegungen zu non-response Problemen erweitert. Unter der Annahme einer Paretoverteilung am oberen Rand der Vermögensverteilung, beläuft sich das geschätzte Gesamtvermögen auf 1,317 Mrd. Euro. Wird im Vergleich dazu, das Gesamtvermögen der österreichischen Haushalte basierend auf den HFCS Daten geschätzt ohne weitere Versuche zu unternehmen für die Untererfassung der Vermögensspitze zu korrigieren, ergibt sich ein Wert von 998 Mrd. Euro. Der Unterschied entspricht einem Anstieg des Durchschnittsvermögens um 81.000 Euro (von 258,000 Euro auf 339,000 Euro). Der Anteil der reichsten 1% der Haushalte am österreichischen Gesamtvermögen steigt dadurch von 25% auf 41%.

Abschließend wird das Aufkommen verschiedener Modelle einer allgemeinen Vermögenssteuer sowohl mit Hilfe der originalen HFCS-Daten als auch mit Hilfe der in Kapitel 5 angepassten Daten geschätzt. Mit der Progression des Steuermodells schwankt das geschätzte Aufkommen zwischen 1,5 Mrd. Euro und 3,6 Mrd. Euro, wenn kein Versuch unternommen wird für die Untererfassung an der Spitze zu kompensieren. Bei Einbeziehung der in Kapitel 5 geschätzten Vermögenswerte schwankt das geschätzte Aufkommen hingegen zwischen 2,9 Mrd. Euro und 8,3 Mrd. Euro je nachdem welcher Steuertarif verwendet wird und welche Annahmen über ein mögliches Ausweichverhalten getroffen werden. Obwohl alle Schätzwerte zusätzlich einer nicht unwesentlichen statistischen Schwankungsbreite unterliegen, zeigen die Ergebnisse, dass eine allgemeine Vermögenssteuer das Potential hat einen erheblichen Anteil zum Steueraufkommen in Österreich beizutragen.

9. Literatur

Atkinson, A. (1995): Incomes and the welfare state. Cambridge, UK: Press Syndicate.

Atkinson, A. (2006): Concentration among the rich, Research Paper, UNU-WIDER, United Nations University (UNU), No. 2006/151, ISBN 9291909351=978-92-9190-935-3

Atkinson, A. (2008): The changing distribution of earnings in OECD countries. Oxford: Oxford Univ. Press.

Atkinson, A. (2014): Inequality: What Can Be Done? Cambridge, Massachusetts: Harvard Univ. Press

Avery, R., Elliehausen, G. and Kennickell, A. (1986): Measuring Wealth with Survey data: An Evaluation of the 1983 Survey of Consumer Finances. Review of Income and Wealth, 34(4), pp.339-369.

Bach, S. und Beznoska, M. (2012): Aufkommens- und Verteilungswirkung einer Wiederbelebung der Vermögenssteuer. DIW: Politikberatung Kompakt 68, Berlin.

Bach, S., Beznoska, M. und Steiner, V. (2012): Aufkommens- und Verteilungswirkung einer Grünen Vermögensabgab. DIW: Politikberatung Kompakt 59, Berlin.

Borgerhoff-Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., Fazzio, I., Gurven, M., Hill, K., Hooper, P., Irons, W., Kaplan, H., Leonetti, D., Low, B., Marlowe, F., McElreath, R., Naidu, S., Nolin, D., Piraino, P., Quinlan, R., Schniter, E., Sear, R., Shenk, M., Smith, E., von Rueden, C. and Wiessner, P. (2009): Intergenerational Wealth Transmission and the Dynamics of Inequality in Small-Scale Societies. Science, 326(5953), pp.682-688.

Clauset, A., Shalizi, C. and Newman, M. (2009): Power-Law Distributions in Empirical Data. SIAM Review, 51(4), pp.661-703.

Cowell, F. (2009): Measuring inequality. Oxford: Oxford University Press.

D'Alessio, G., Faiella, I. (2002): Non-response behaviour in the Bank of Italy's Survey of Household Income and Wealth, No 462, Temi di discussione (Economic working papers), Bank of Italy, Economic Research and International Relations Area.

Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B., Springholz, F. and Wildauer, R. (2016): Correcting for the Missing Rich: An Application to Wealth Survey Data. Review of Income and Wealth, 62(4), pp.605-627.

Eckerstorfer, P., Halak, J., Kapeller, J., Schütz, B. and Wildauer, R. (2013): Bestände und Verteilung der Vermögen in Österreich. Materialien zu Wirtschaft und Gesellschaft, 122, pp.1-49.

European Central Bank (2013): The Eurosystem Household Finance and Consumption Survey. Methodological Report for the first wave. Statistics Paper Series. No 1/2013.

European Central Bank (2016): The Household Finance and Consumption Survey: methodological report for the second wave. Statistics Paper Series. No 17/2016.

Gabaix, X. (2016): Power Laws in Economics: An Introduction. Journal of Economic Perspectives, 30(1), pp. 185-206.

Gibrat, R. (1931): Les Inégalités économiques, applications. Paris: Libr. du Recueil Sirey.

Guttmann, R. and Plihon, D. (2010): Consumer debt and financial fragility. International Review of Applied Economics, 24(3), pp.269-283.

IMF (2015): Causes and Consequences of Income Inequality: A Global Perspective. IMF Staff Discussion Note.

Kennickell, A.B., 2005. The Good Shepherd: Sample Design and Control for Wealth Measurement in the Survey of Consumer Finances. Luxembourg Wealth Study Conference.

Kennickell, A.B., 2008. The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances. Irving Fisher Committee Bulletin, (28), pp.403–408.

Kennickell, A. B. (2010): Try, Try Again: Response and Nonresponse in the 2009 SCF Panel. Proceedings of the Section on Survey Research Methods.

Kennickell, A. und McManus, D. (1993): Sampling for Household Financial Characteristics Using Frame Information on Past Income. Paper presented at the 1993 Joint Statistical Meetings, Atlanta.

Klass, O., Biham, O., Levy, M., Malcai, O. and Solomon, S. (2006): The Forbes 400 and the Pareto wealth distribution. Economics Letters, 90(2), pp.290-295.

Kratz, M. and Resnick, S. I. (1996): The qq-estimator and heavy tails. Communications in Statistics. Stochastic Models, 12(4), pp.699-724.

Little, R. J. A. and Rubin, D. B. (2002): Statistical Analysis with Missing Data, Second Edition, Hoboken, N.J: John Wiley.

Milanovic, B. (2005): Worlds Apart. Measuring International and Global Inequality. New Jersey: Princeton University Press.

Milanovic, B. (2016): Income inequality is cyclical. Nature, 537(7621), pp.479-482.

Newman, M. (2005): Power Laws, Pareto Distributions and Zipf's law. Contemporary Physics, 46, 323-351.

Österreichische Nationalbank (2016a): Household Finance and Consumption Survey des Eurosystems 2014: Erste Ergebnisse für Österreich (zweite Welle). Available at: https://www.hfcs.at/dam/jcr:f1c6641e-f691-426b-a690-1da5cbf0203d/HFCS%20Erste%20Ergebnisse%20Juni_16-screen.pdf [accessed: 28 October 2017].

Österreichische Nationalbank (2016b): Household Finance and Consumption Survey des Eurosystems 2014: Methodische Grundlagen für Österreich (zweite Welle). Available at: https://www.hfcs.at/dam/jcr:13f31fd0-af99-4dce-980d-7fbd1f24ac80/HFCS_Methodische%20_Grundlagen_2016.pdf [accessed: 28 October 2017]

Osier, G. (2016): Unit non-response in household wealth surveys. Experience from the Eurosystem's Household Finance and Consumption Survey. Statistics Paper Series. No 15/2016.

Piketty, T. (2014): Capital in the twenty-first century. Cambridge, Massachusetts: Harvard University Press.

Rigney, D. (2010): The Matthew effect. New York: Columbia University Press.

Rubin, D. (1987): Multiple imputation for nonresponse in surveys. Hoboken, N.J: John Wiley.

Stiglitz, J. (2012): The Price of Inequality: How Today's Divided Society Endangers Our Future. New York: W.W. Norton.

Vermeulen, P. (2014): How fat is the top tail of the wealth distribution? ECB-Working Paper Series, (1692).

Vermeulen, P. (2016): Estimating the Top Tail of the Wealth Distribution. American Economic Review, 106(5), pp.646-650.

Wald, A. (1945): Statistical Decision Functions Which Minimize the Maximum Risk. Annals of Mathematics, 46, 265–80, 1945.

Anhang I: Perzentillisten

Perzentillisten HFCS Daten für Österreich Welle II - Originaldaten

Anhang I: Perzentilliste auf Basis der HFCS-Daten					
Perz-entil	Gesamtvermögen im Perzentil	Durchschnittsvermögen im Perzentil	Perz-entil	Gesamtvermögen im Perzentil	Durchschnittsvermögen im Perzentil
1	-€ 3,787,401,027	-€ 96,561	51	€ 3,488,840,666	€ 89,834
2	-€ 503,697,374	-€ 13,075	52	€ 3,743,903,811	€ 97,676
3	-€ 190,048,085	-€ 4,932	53	€ 4,142,765,587	€ 107,049
4	-€ 87,532,050	-€ 2,269	54	€ 4,411,020,128	€ 113,682
5	-€ 29,664,503	-€ 762	55	€ 4,658,169,527	€ 119,376
6	-€ 3,430,573	-€ 89	56	€ 4,825,874,055	€ 126,725
7	€ 1,073,866	€ 28	57	€ 5,200,897,105	€ 134,112
8	€ 7,058,532	€ 181	58	€ 5,480,122,522	€ 142,610
9	€ 15,729,281	€ 410	59	€ 5,811,916,258	€ 151,691
10	€ 30,492,687	€ 789	60	€ 6,200,283,409	€ 159,158
11	€ 45,128,042	€ 1,168	61	€ 6,497,017,739	€ 166,385
12	€ 59,406,122	€ 1,537	62	€ 6,658,810,061	€ 174,805
13	€ 79,588,539	€ 2,057	63	€ 7,143,591,542	€ 184,945
14	€ 102,330,684	€ 2,673	64	€ 7,481,844,654	€ 194,309
15	€ 124,906,584	€ 3,222	65	€ 7,936,041,052	€ 204,116
16	€ 146,886,644	€ 3,779	66	€ 8,196,641,781	€ 213,811
17	€ 177,165,980	€ 4,492	67	€ 8,569,524,321	€ 220,946
18	€ 187,918,385	€ 5,006	68	€ 8,828,957,414	€ 227,614
19	€ 214,013,866	€ 5,488	69	€ 9,177,909,101	€ 236,390
20	€ 237,343,040	€ 6,113	70	€ 9,418,631,589	€ 245,950
21	€ 260,895,940	€ 6,773	71	€ 10,057,739,718	€ 257,899
22	€ 290,579,101	€ 7,577	72	€ 10,180,759,345	€ 266,945
23	€ 327,548,215	€ 8,327	73	€ 10,838,029,910	€ 277,805
24	€ 347,017,176	€ 9,170	74	€ 11,163,332,037	€ 287,480
25	€ 393,123,106	€ 10,118	75	€ 11,509,532,209	€ 299,254
26	€ 424,547,501	€ 11,090	76	€ 11,977,553,750	€ 310,954
27	€ 467,357,549	€ 11,972	77	€ 12,461,250,019	€ 321,663
28	€ 499,316,355	€ 12,922	78	€ 12,908,027,006	€ 332,936
29	€ 525,271,188	€ 13,782	79	€ 13,410,532,081	€ 346,620
30	€ 585,552,540	€ 15,113	80	€ 13,898,428,180	€ 358,715
31	€ 640,074,569	€ 16,480	81	€ 13,998,812,856	€ 367,862
32	€ 683,566,654	€ 17,783	82	€ 14,564,554,675	€ 377,204
33	€ 742,617,526	€ 19,097	83	€ 15,107,896,467	€ 390,546
34	€ 804,917,152	€ 20,761	84	€ 15,693,452,725	€ 403,881
35	€ 845,538,824	€ 22,209	85	€ 15,831,633,331	€ 415,075
36	€ 941,163,812	€ 23,980	86	€ 16,651,925,448	€ 429,947
37	€ 984,374,690	€ 25,772	87	€ 17,323,223,457	€ 445,756
38	€ 1,083,700,907	€ 27,727	88	€ 17,735,236,100	€ 461,972
39	€ 1,156,871,383	€ 30,208	89	€ 18,663,337,208	€ 482,531
40	€ 1,268,650,822	€ 32,811	90	€ 19,473,721,157	€ 506,975
41	€ 1,398,513,811	€ 35,973	91	€ 20,822,789,219	€ 529,606
42	€ 1,508,494,828	€ 39,264	92	€ 21,274,388,242	€ 562,269
43	€ 1,651,710,513	€ 42,763	93	€ 23,273,639,040	€ 604,246
44	€ 1,823,925,447	€ 47,302	94	€ 25,703,410,726	€ 658,576
45	€ 2,017,047,419	€ 52,001	95	€ 29,445,091,186	€ 759,053
46	€ 2,199,680,800	€ 57,482	96	€ 32,969,286,315	€ 847,449
47	€ 2,469,761,921	€ 63,566	97	€ 37,378,504,791	€ 980,399
48	€ 2,664,572,135	€ 68,861	98	€ 47,125,664,840	€ 1,218,196
49	€ 2,896,553,981	€ 75,121	99	€ 62,361,063,542	€ 1,618,187
50	€ 3,201,209,599	€ 82,753	100	€ 254,522,764,362	€ 6,703,743
		Gesamtvermögen:		€ 998,129,766,372	

Anmerkung: Eine Besonderheit von imputierten Datensätzen ist, dass die daraus gewonnenen Resultate immer nur Durchschnittswerte über alle Imputationen sein können. Nach Little und Rubin (2002) können solche Ergebnisse deshalb nicht als Grundlage für weitere Berechnungen verwendet werden. Weiters ist bei dieser detaillierten Darstellung Vorsicht geboten, weil die Basis für jedes Perzentil nur jeweils eine kleine Menge von Beobachtungen bildet und die Standardfehler daher hoch sind.

Perzentillisten HFCS Daten für Österreich Welle II - Schätzung

Perz-entil	Gesamtvermögen im Perzentil	Durchschnittsvermögen im Perzentil	Perz-entil	Gesamtvermögen im Perzentil	Durchschnittsvermögen im Perzentil
1	-€ 3,771,561,163	-€ 96,019	51	€ 3,560,066,218	€ 91,720
2	-€ 503,022,002	-€ 12,950	52	€ 3,857,449,872	€ 100,207
3	-€ 185,824,907	-€ 4,843	53	€ 4,235,905,558	€ 109,028
4	-€ 85,936,613	-€ 2,223	54	€ 4,442,034,111	€ 115,236
5	-€ 28,383,871	-€ 735	55	€ 4,617,116,361	€ 121,209
6	-€ 3,157,333	-€ 81	56	€ 5,001,807,189	€ 128,618
7	€ 1,173,836	€ 31	57	€ 5,365,410,284	€ 136,467
8	€ 7,288,726	€ 188	58	€ 5,514,443,960	€ 145,436
9	€ 16,443,179	€ 423	59	€ 6,075,169,556	€ 154,078
10	€ 32,081,810	€ 816	60	€ 6,228,346,429	€ 161,559
11	€ 44,975,055	€ 1,194	61	€ 6,428,395,248	€ 168,802
12	€ 60,680,512	€ 1,561	62	€ 6,856,298,200	€ 177,820
13	€ 80,585,605	€ 2,098	63	€ 7,346,280,458	€ 188,075
14	€ 106,424,469	€ 2,720	64	€ 7,575,791,688	€ 197,552
15	€ 126,481,732	€ 3,268	65	€ 8,039,350,054	€ 207,513
16	€ 144,920,220	€ 3,834	66	€ 8,266,605,956	€ 216,347
17	€ 176,961,451	€ 4,539	67	€ 8,767,883,777	€ 223,450
18	€ 196,307,782	€ 5,042	68	€ 8,840,831,097	€ 230,093
19	€ 213,263,465	€ 5,555	69	€ 9,267,317,985	€ 239,572
20	€ 235,100,101	€ 6,166	70	€ 9,681,795,044	€ 250,742
21	€ 265,944,886	€ 6,839	71	€ 10,108,188,304	€ 260,915
22	€ 297,602,893	€ 7,661	72	€ 10,424,316,997	€ 271,183
23	€ 328,985,073	€ 8,418	73	€ 11,009,761,456	€ 281,362
24	€ 355,404,189	€ 9,290	74	€ 11,179,208,785	€ 291,736
25	€ 392,650,638	€ 10,242	75	€ 11,720,233,982	€ 303,973
26	€ 432,670,784	€ 11,205	76	€ 12,173,856,247	€ 315,229
27	€ 473,729,991	€ 12,089	77	€ 12,526,599,059	€ 325,877
28	€ 509,128,965	€ 13,078	78	€ 12,929,366,299	€ 337,903
29	€ 526,071,520	€ 13,946	79	€ 13,657,145,446	€ 351,585
30	€ 591,741,882	€ 15,332	80	€ 14,278,028,144	€ 363,073
31	€ 652,337,679	€ 16,691	81	€ 14,293,220,399	€ 371,322
32	€ 697,302,672	€ 17,994	82	€ 14,708,687,181	€ 382,482
33	€ 736,445,863	€ 19,369	83	€ 15,237,202,291	€ 396,816
34	€ 812,812,930	€ 21,035	84	€ 15,816,940,574	€ 408,623
35	€ 871,478,212	€ 22,470	85	€ 16,206,325,414	€ 420,768
36	€ 944,476,325	€ 24,309	86	€ 16,989,232,476	€ 437,121
37	€ 998,419,210	€ 26,118	87	€ 17,516,000,903	€ 453,049
38	€ 1,096,730,654	€ 28,138	88	€ 18,040,104,545	€ 471,017
39	€ 1,180,661,726	€ 30,701	89	€ 18,898,075,093	€ 493,422
40	€ 1,288,714,585	€ 33,410	90	€ 20,073,417,584	€ 517,036
41	€ 1,412,621,336	€ 36,652	91	€ 21,106,279,712	€ 544,031
42	€ 1,545,416,920	€ 39,924	92	€ 22,728,844,332	€ 581,599
43	€ 1,716,570,491	€ 43,710	93	€ 23,793,644,803	€ 626,885
44	€ 1,831,090,018	€ 48,338	94	€ 27,179,055,491	€ 702,909
45	€ 2,092,799,287	€ 53,194	95	€ 31,142,721,475	€ 801,905
46	€ 2,238,387,481	€ 58,967	96	€ 35,184,353,525	€ 904,206
47	€ 2,528,326,745	€ 64,870	97	€ 41,566,792,448	€ 1,074,065
48	€ 2,731,262,419	€ 70,386	98	€ 53,533,086,856	€ 1,390,025
49	€ 2,937,498,733	€ 76,917	99	€ 76,892,240,929	€ 2,013,261
50	€ 3,259,725,565	€ 84,542	100	€ 533,985,842,784	€ 14,045,856
		Gesamtvermögen:		€ 1,317,478,884,304	

Anmerkung: Eine Besonderheit von imputierten Datensätzen ist, dass die daraus gewonnenen Resultate immer nur Durchschnittswerte über alle Imputationen sein können. Nach Little und Rubin (2002) können solche Ergebnisse deshalb nicht als Grundlage für weitere Berechnungen verwendet werden. Weiters ist bei dieser detaillierten Darstellung Vorsicht geboten, weil die Basis für jedes Perzentil nur jeweils eine kleine Menge von Beobachtungen bildet und die Standardfehler daher hoch sind.

Anhang II: Schätzung der Paretoverteilung und Korrektur der Daten (Mathematica Code)

Schätzung der Verteilungsfunktion und Modifikation der Daten

Vorbereitende Tätigkeiten

Einlesen der Daten

Hier werden der Daten der zweiten HFCS Welle eingelesen wobei "*wealth*" die Nettovermögenswerte indiziert und "*weight*" die entsprechende Gewichtung der jeweiligen Beobachtung. Der Zusatz "2" verweist auf den Ursprung der Daten in der zweiten Welle der HFCS. Unter "*richlist*" wird eine Reichenliste eingelesen, die auf Basis der Publikationen des Trend erstellt worden ist. Die durch das Hinzufügen dieser Reichenliste erweiterten Datenvektoren werden wiederum durch "*wealth2r*" und "*weight2r*" bezeichnet.

```
SetDirectory["YOUR PATH HERE"];
```

In[2]:=
```
wealth2 = Table[ Import["wave2.csv"][[k]], {k, 1, 5}];
weight2 = Table[Import["wave2.csv"][[k + 5]], {k, 1, 5}];
richlist2 = Transpose[Import["richlist-waveII.xlsx"][[1]]][[1]];
wealth2r = Table[richlist2 ~Join~ wealth2[[k]], {k, 1, 5}];
weight2r = Table[Table[1, {Length[richlist2]}] ~Join~ weight2[[k]], {k, 1, 5}];
```

Definition von Funktion

- **Cramer von Mises Test**

Hier wird eine manuelle Programmierung des Cramer von Mises Tests vorgenommen, dessen Funktion und Ergebnis weiter unten genauer besprochen werden.

In[7]:=
```
cvm[data_List, percposition_, gew_List, paretoalpha_, percvalue_] := Module[{},
  total = Plus @@ gew[[1 ;; percposition]];
  summe = 0;
  summe = summe + 1 / (12 * Plus @@ gew[[1 ;; percposition]]);
  a = 1;
  index = 1;
  b = Reverse[gew[[1 ;; percposition]]][[1]];
  d = 0;
  While[
   b < total, summe = summe - 1/(12 total^2) (-1 + a - b)
       (3 - 8 a + 4 a^2 - 4 b + 4 a b + 4 b^2 + 12 d total - 12 a d total - 12 b d total + 12 d^2 total^2);
   a = b + 1;
   b = b + Reverse[gew[[1 ;; percposition]]][[index + 1]];
   index++;
   d = 1 - (percvalue / Reverse[data[[1 ;; percposition]]][[index]]) ^paretoalpha];
  summe / Mean[gew[[1 ;; percposition]]]]
```

Schätzung der Verteilungsfunktion

Perzentilberechnung

Im Folgenden werden die Perzentilgrenzen und die damit korrespondierenden Vermögenswerte berechnet.

In[8]:=
```
percpositions2 =
  Table[Length[Select[Accumulate[weight2r[[k]]] / Total[weight2r[[k]]], # < 0.01 * n &]],
   {k, 1, 5}, {n, 1, 100}];
percvalues2 = Table[wealth2r[[k]][[Length[Select[Accumulate[weight2r[[k]]] /
      Total[weight2r[[k]]], # < 0.01 * n &]]]], {k, 1, 5} , {n, 1, 100}];
```

Schätzung der Paretoverteilungen

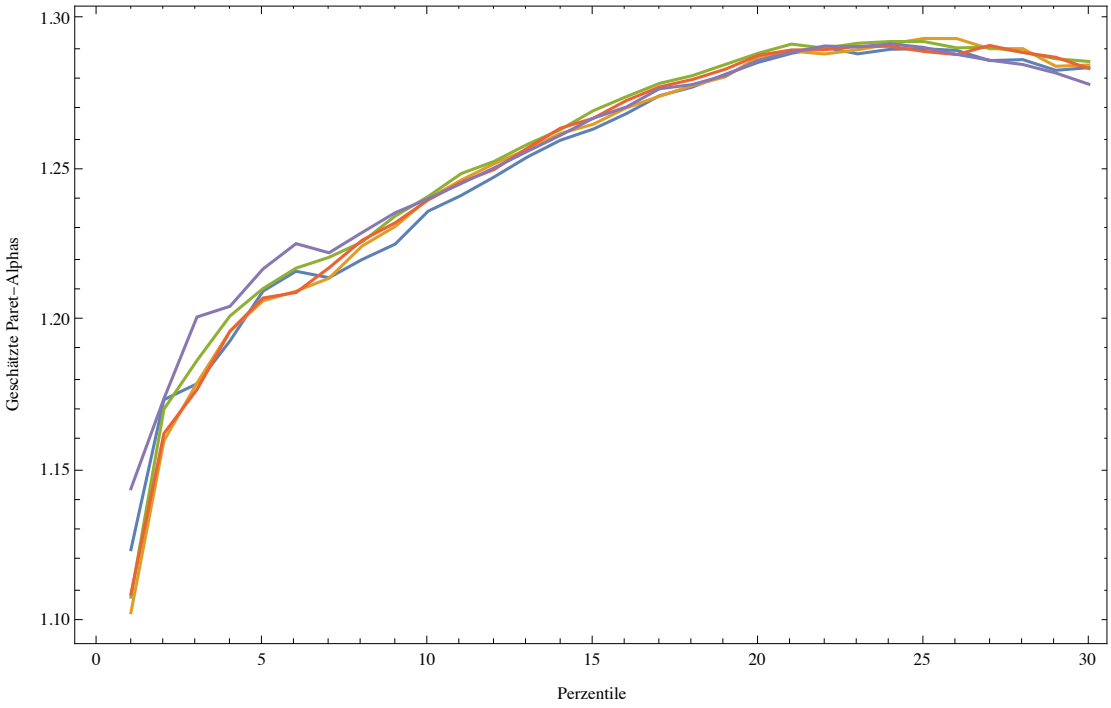
■ Schätzung der Verteilung über alle Implicates

Hier werden Paretoverteilungen für die obersten 30 Perzentile über alle 5 Implicates der Daten erstellt, sowie deren Pareto-Alphas gesammelt. Die statistische Plausibilität dieser Verteilungen wird anhand eines Cramer von Mises Tests beurteilt, dessen Teststatistiken anzeigen wie gut die Verteilung der Daten mit einer hypothetischen Verteilung korrespondieren. Die so gesammelten Teststatistiken und Steigungsparameter werden unten bildlich dargestellt.

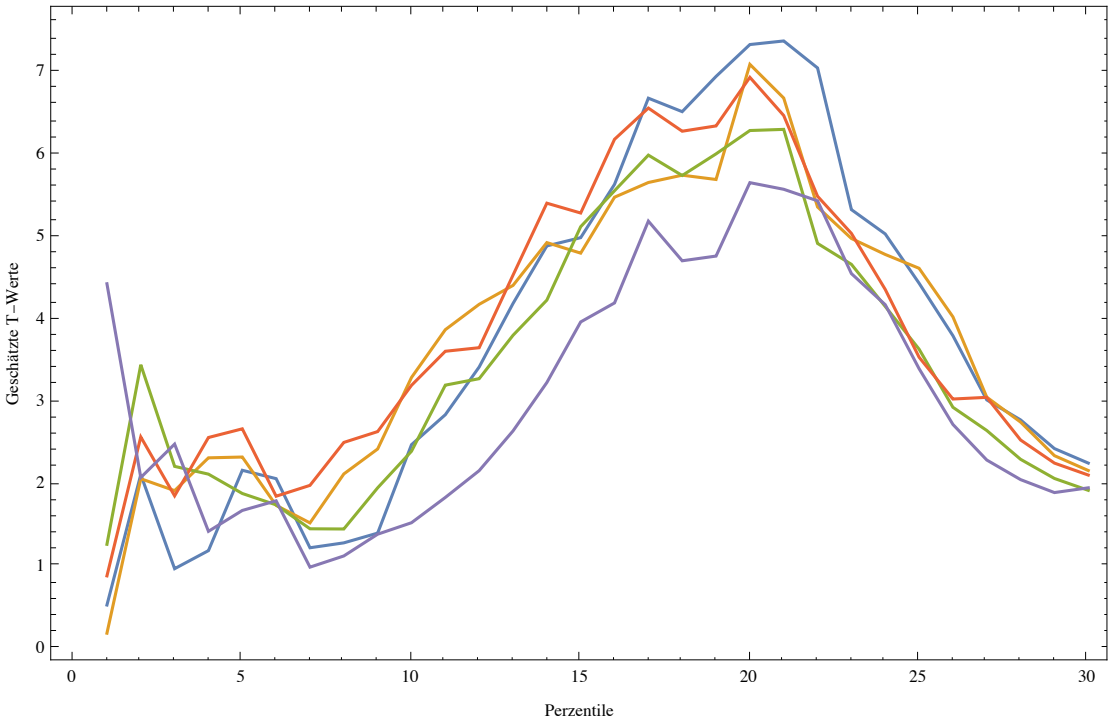
In[10]:=
```
alpha2 = Table[data = Transpose[{
      Log[wealth2r[[k, 1 ;; percpositions2[[k, p]]]] /
        Min[wealth2r[[k, 1 ;; percpositions2[[k, p]]]]]],
      Log[Accumulate[weight2r[[k, 1 ;; percpositions2[[k, p]]]]] /
        Total[weight2r[[k, 1 ;; percpositions2[[k, p]]]]]]
     }];
   LinearModelFit[data, {alpha}, {alpha}, IncludeConstantBasis → False][[1]][[2]][[1]] *
    (-1), {k, 1, 5}, {p, 1, 30}];
ed = Table[ParetoDistribution[percvalues2[[k, p]], alpha2[[k, p]]],
   {k, 1, 5}, {p, 1, 30}];
teststat2 = Table[cvm[wealth2r[[k]], percpositions2[[k, n]], weight2r[[k]],
    alpha2[[k, n]], percvalues2[[k, n]]], {k, 1, 5}, {n, 1, 30}];
```

In[13]:=
```
alphaplotkorr2 = ListPlot[alpha2, Joined → True, Frame → True, FrameLabel →
    {{"Geschätzte Paret-Alphas", None}, {"Perzentile", None}}, ImageSize → Large]
tstatplotkorr2 = ListPlot[teststat2, Joined → True, Frame → True,
  FrameLabel → {{"Geschätzte T-Werte", None}, {"Perzentile", None}}, ImageSize → Large]
```

Out[13]=



Out[14]=



■ Finden des MaxiMin - Perzentils der T-Statistiken

Hier werden die Minima und Maxima der oben errechneten Teststatistiken im jeweiligen Perzentil und über alle Implicates verglichen. Als Basis für die spätere Schätzung und Datenmodifikation wird daraus ein optimaler Ansatzpunkt der Paretoverteilung gewählt. Spezifisch wird der Punkt gesucht an dem die Minima der Teststatistiken über alle Implicates maximal sind. Es ergibt sich dabei das 7te Perzentil von oben, sprich das 93te Perzentil, als optimaler Ansatz der geschätzten Paretoverteilung. Ebenso wird die durchschnittliche Teststatistik des CVM an diesem Punkt ausgelesen.

```
In[15]:= cutoffpoint2 = Flatten[Position[Transpose[teststat2],
             Min[Table[Max[Transpose[teststat2][[p]]], {p, 1, 30}]]][[1]]
         Mean[teststat2][[cutoffpoint2]]
Out[15]= 7
Out[16]= 1.43635
```

- **Wert des Pareto-Alphas an der Stellle des 93. Perzentils**

```
In[17]:= finalalphas2 = Transpose[alpha2][[cutoffpoint2]]
         finalalpha2 = Mean[Transpose[alpha2][[cutoffpoint2]]]
Out[17]= {1.21408, 1.21393, 1.22088, 1.2174, 1.22237}
Out[18]= 1.21773
```

Modifikation der Daten

Parameterdefinition

Untenstehend finden sich Variablennotationen für die Modifikation der Daten und die Formulierung einer Ergebnistabelle deren Bestandteile unter "*varlable*" gelistet sind,

```
In[19]:= implabel = {"Implicate 1", "Implicate 2", "Implicate 3", "Implicate 4", "Implicate 5"};
         varlabel = {"Number of Obs\n > 4 Mio.",
            "Number of HH\n > 4 Mio.", "Number of Obs\n between cutoffpoint and 4 Mio.",
            "Number of HH\n between cutoffpoint and 4 Mio.",
            "Number of HH > 4 Mio. \n according to Pareto Dist.",
            "Adjustment parameter\n for Sampling Weigths", "Growth of HH > 4 Mio."};
         upperthreshold = 4 * 10^6;
         results = Table[0, {Length[varlabel]}, {Length[implabel]}];
```

Modifizierte Datenstruktur: Beobachtungen

Der untenstehende Prozess folgt folgenden Zwischenschritten. Zuerst wird, angeührt an zweiter Stelle [[2]] der Ergebnistabelle "*results*" die Anzahl der Beobachtungen aus dem originalen HFCS II Datensatz errechnet die eine Vermögensgrenze von 4mio. überschreiten. Ab diesem "*cutoffvalue*" werden Haushalte im Zuge der Modifikation neu generiert. An dritter Stelle [[3]] der Ergebnistabelle werden diese Beobachtungen gewichtet und die Anzahl der dadurch repräsentierten Haushalte dargestellt. Stelle [[4]] berechnet dann die Anzahl der Beobachtungen zwischen dem oben bestimmten Startpunkt der Paretoverteilung, dem "*cutoffpoint*" und dem "*cutoffvalue*", Stelle [[5]] wiederum die Anzahl der dadurch repräsentierten Haushalte. An Stelle [[6]] werden auf Basis der oben geschätzten Parameter der Paretoverteilung jene Haushalte berechnet die über dem cutoffvalue liegen. Da sich die "*sample size*" verändert, werden in Rechenschritt [[7]] die Parameter einer Anpassung der Haushaltsgewichte errechnet. Die Änderung der Anzahl an Haushalten ist in der Ergebnistabelle an Stelle [[8]] angeführt.

```
In[23]:= results[[1]] = Table[Length[Select[wealth2[[k]], # ≥ upperthreshold &]], {k, 1, 5}]
         results[[2]] = Table[Total[Take[weight2[[k]], results[[1, k]]]], {k, 1, 5}]
         results[[3]] = Table[Length[Select[wealth2[[k]],
             upperthreshold ≥ # ≥ percvalues2[[k, cutoffpoint2]] &]], {k, 1, 5}]
         results[[4]] = Table[Total[Drop[Drop[weight2[[k]], results[[1, k]]],
             -Length[Select[wealth2[[k]], # < percvalues2[[k, cutoffpoint2]] &]]]], {k, 1, 5}]
         results[[5]] = Table[Round[results[[4, k]] *
             ((1 - CDF[ParetoDistribution[percvalues2[[k, cutoffpoint2]], finalalphas2[[k]]],
                   upperthreshold]) / (1 - (1 - CDF[ParetoDistribution[percvalues2[[k,
                       cutoffpoint2]], finalalphas2[[k]]], upperthreshold])))], {k, 1, 5}]
         results[[6]] = Table[(results[[5, k]] - results[[2, k]]) /
           (Total[weight2[[k]]] - (results[[2, k]] + results[[4, k]])),
           {k, 1, 5}]
         results[[7]] = Table[results[[5, k]] - results[[2, k]], {k, 1, 5}]
Out[23]= {11, 9, 7, 7, 6}
Out[24]= {16 034.5, 13 667.5, 9117.78, 11 391.8, 8195.38}
Out[25]= {187, 185, 186, 190, 187}
Out[26]= {253 929., 254 264., 260 819., 258 754., 261 862.}
Out[27]= {29 516, 29 357, 30 557, 29 663, 31 376}
Out[28]= {0.00375263, 0.00436474, 0.00596762, 0.00508609, 0.00645256}
Out[29]= {13 481.5, 15 689.5, 21 439.2, 18 271.2, 23 180.6}
```

Tabelle der Zwischenergebnisse

```
In[30]:= Prepend[Transpose[Prepend[results, implabel]],
           {"Implicate"} ~ Join ~ varlabel] // TableForm
```

Modifikation Datenstruktur: Gewichte

Hier werden die Gewichte der Haushalte nun abgeschmolzen um der veränderten Haushaltsanzahl nachzukommen, sowie die Summen der Gewichtsvektoren durch Differenzenbildung kontrolliert um die Gewichtsvektorskonstruktion zu überprüfen.

```
In[31]:= part1 = Table[Table[1, {results[[5, k]]}], {k, 1, 5}];
         part2 = Table[Take[Drop[weight2[[k]], results[[1, k]]], results[[3, k]]], {k, 1, 5}];
         part3 = Table[Drop[weight2[[k]], (results[[1, k]] + results[[3, k]])] *
             (1 - results[[6, k]]), {k, 1, 5}];
         corrweight2 = Table[part1[[k]] ~ Join ~ part2[[k]] ~ Join ~ part3[[k]], {k, 1, 5}];
In[35]:= Table[0 == Round[Total[weight2[[k]]] - Total[corrweight2[[k]]]], {k, 1, 5}]
Out[35]= {True, True, True, True, True}
```

Modifikation: Generierung neuer Haushalte

Auf Basis der obigen Berechnungen und Zwischenergebnisse können nun entsprechende neue Haushalte generiert werden, die, in neue Datenvektoren gefasst, eine solidere Basis für die Schätzung österreichischer Vermögensbestände liefert als die von der HFCS II präsentierten Daten.

In[36]:=
```
corrwealth2 =
 Table[Table[If[((results[[5, k]] + results[[4, k]]) / j)^(1 / finalalphas2[[k]]) *
       percvalues2[[k, cutoffpoint2]] > 1 000 000 000, 1 000 000 000,
     ((results[[5, k]] + results[[4, k]]) / j)^(1 / finalalphas2[[k]]) *
      percvalues2[[k, cutoffpoint2]]], {j, 1, results[[5, k]]}] ~
   Join ~ Drop[wealth2[[k]], results[[1, k]]], {k, 1, 5}]
```

Out[36]=
```
{{1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000,
  1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000,
  1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000,
  1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000,
  1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000, 1 000 000 000,
  ··· 32 442 ···, -21 350, -21 768, -21 850, -22 464, -23 633, -23 698, -26 466,
  -26 981, -30 686, -34 000, -34 994, -35 462, -35 820, -37 895, -47 500, -47 800,
  -49 204, -60 503, -63 400, -65 199, -67 900, -70 000, -103 647, -146 554,
  -161 963, -197 497, -233 000, -236 000, -375 840, -392 126}, ··· 3 ···, {··· 1 ···}}
```

large output | show less | show more | show all | set size limit...

Vergleich, Finalisierung und Export

Berechnung von Perzentillisten

Hier werden Perzentillisten sowohl für die Originaldaten und geschätzten Daten errechnet, als auch eine Schätzung des Gesamtvermögens vorgenommen.

■ Originaldaten

In[37]:=
```
pweights = Table[Accumulate[weight2[[k]]] / Total[weight2[[k]]] * 100, {k, 1, 5}];
borders = Table[Length[Select[pweights[[l]], # ≤ k &]], {l, 1, 5}, {k, 1, 100}];
perctotal2 =
  Mean[Table[Total[wealth2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]] *
      weight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]],
    {l, 1, 5}, {k, 1, 100}]];
percaver2 = Mean[Table[Total[wealth2[[l,
        If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]] *
      weight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]] /
    Total[weight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]], {l,
    1, 5}, {k, 1, 100}]];
limits2 = Mean[Table[wealth2[[l, borders[[l, k]] + 1]], {l, 1, 5}, {k, 1, 99}]];
totalwealth2 = Mean[Total[Transpose[Table[weight2[[k]] * wealth2[[k]], {k, 1, 5}]]]]
(*auxiliar definitions for shares*)
totwealth2imp = Total[Transpose[Table[weight2[[k]] * wealth2[[k]], {k, 1, 5}]]];
perctotal2imp = Table[Total[
    wealth2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]] * weight2[[l,
      If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]], {l, 1, 5}, {k, 1, 100}];
```

Out[42]=
$$9.9813 \times 10^{11}$$

■ Geschätzte Daten

In[45]:=
```
pweights =
  Table[Accumulate[corrweight2[[k]]] / Total[corrweight2[[k]]] * 100, {k, 1, 5}];
borders = Table[Length[Select[pweights[[l]], # ≤ k &]], {l, 1, 5}, {k, 1, 100}];
perctotalc2 = Mean[
   Table[Total[corrwealth2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]] *
      corrweight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]],
    {l, 1, 5}, {k, 1, 100}]];
percaverc2 = Mean[Table[Total[corrwealth2[[l,
        If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]] *
      corrweight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]] /
    Total[corrweight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]],
   {l, 1, 5}, {k, 1, 100}]];
limitsc2 = Mean[Table[corrwealth2[[l, borders[[l, k]] + 1]], {l, 1, 5}, {k, 1, 99}]];
totalwealthc2 = Mean[Table[Total[corrweight2[[k]] * corrwealth2[[k]]], {k, 1, 5}]]
(*auxiliar definitions for shares*)
totwealthc2imp = Table[Total[corrweight2[[k]] * corrwealth2[[k]]], {k, 1, 5}];
perctotalc2imp =
  Table[Total[corrwealth2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]] *
     corrweight2[[l, If[k == 1, 1, borders[[l, k - 1]] + 1] ;; borders[[l, k]]]]],
   {l, 1, 5}, {k, 1, 100}];
```

Out[50]= 1.31684×10^{12}

Vergleich der Ergebnisse und Export

Untenstehenden findet sich eine Auswahl von Vergleichswerten zu den Originaldaten sowie den geschätzten Daten. Die “2” am Ende der Ergebnisvariablen gibt die Welle “2” der HFCS an und das “c” nach der Rechengrößenbeschreibung (“totwealth”=Nettogesamtvermögen etc.) verweist auf die Verwendung der gerschätzten Daten. Weiter unten finden sich Berechnungen von Verteilungsanteilen und schließlich werden die finalen Ergebnisse für den Export aufbereitet.

■ Durchschnitts- und Gesamtvermögen

In[53]:=
```
totalwealth2 = Mean[Total[Transpose[Table[weight2[[k]] * wealth2[[k]], {k, 1, 5}]]]]
avwealth2 = Mean[Total[Transpose[Table[weight2[[k]] * wealth2[[k]], {k, 1, 5}]]]] /
  Total[weight2[[1]]]
totalwealthc2 = Mean[Table[Total[corrweight2[[k]] * corrwealth2[[k]]], {k, 1, 5}]]
avwealthc2 =
 Mean[Table[Total[corrweight2[[k]] * corrwealth2[[k]]], {k, 1, 5}]] / Total[weight2[[1]]]
```

Out[53]= 9.9813×10^{11}

Out[54]= 258 414.

Out[55]= 1.31684×10^{12}

Out[56]= 340 927.

Vermögensanteile

Top1 %

```
In[57]:= sharetop1AGG = perctotal2[[1]] / totalwealth2
         sharetop1cAGG = perctotalc2[[1]] / totalwealthc2
         sharetop1 = Mean[perctotal2imp[[All, 1]] / totwealth2imp]
         sharetop1c = Mean[perctotalc2imp[[All, 1]] / totwealthc2imp]
```

```
Out[57]= 0.255
Out[58]= 0.405017
Out[59]= 0.254055
Out[60]= 0.405044
```

Top5 %

```
In[61]:= sharetop5AGG = Total[perctotal2[[1 ;; 5]]] / totalwealth2
         sharetop5cAGG = Total[perctotalc2[[1 ;; 5]]] / totalwealthc2
         sharetop5 = Mean[Total[Transpose[perctotal2imp[[All, 1 ;; 5]]]] / totwealth2imp]
         sharetop5c = Mean[Total[Transpose[perctotalc2imp[[All, 1 ;; 5]]]] / totwealthc2imp]
```

```
Out[61]= 0.435171
Out[62]= 0.562502
Out[63]= 0.434437
Out[64]= 0.562413
```

Top10 %

```
In[65]:= sharetop10AGG = Total[perctotal2[[1 ;; 10]]] / totalwealth2
         sharetop10cAGG = Total[perctotalc2[[1 ;; 10]]] / totalwealthc2
         sharetop10 = Mean[Total[Transpose[perctotal2imp[[All, 1 ;; 10]]]] / totwealth2imp]
         sharetop10c = Mean[Total[Transpose[perctotalc2imp[[All, 1 ;; 10]]]] / totwealthc2imp]
```

```
Out[65]= 0.555916
Out[66]= 0.657991
Out[67]= 0.555346
Out[68]= 0.657897
```

Top20 %

```
In[69]:= sharetop20AGG = Total[perctotal2[[1 ;; 20]]] / totalwealth2
         sharetop20cAGG = Total[perctotalc2[[1 ;; 20]]] / totalwealthc2
         sharetop20 = Mean[Total[Transpose[perctotal2imp[[All, 1 ;; 20]]]] / totwealth2imp]
         sharetop20c = Mean[Total[Transpose[perctotalc2imp[[All, 1 ;; 20]]]] / totwealthc2imp]
```

```
Out[69]= 0.721269
Out[70]= 0.785403
Out[71]= 0.720919
Out[72]= 0.785339
```

Lower50 %

In[73]:=
```
sharelower50AGG = Total[perctotal2[[51 ;; 100]]] / totalwealth2
sharelower50cAGG = Total[perctotalc2[[51 ;; 100]]] / totalwealthc2
sharelower50 = Mean[Total[Transpose[perctotal2imp[[All, 51 ;; 100]]]] / totwealth2imp]
sharelower50c = Mean[Total[Transpose[perctotalc2imp[[All, 51 ;; 100]]]] / totwealthc2imp]
```

Out[73]= 0.0319913

Out[74]= 0.0247655

Out[75]= 0.0320237

Out[76]= 0.0247785

Export

Export der finalen Ergebnisse

In[77]:=
```
finalresults =
  {{"Durchschnitt", "Gesamt", "Anteil Top1%", "Anteil Top5%", "Anteil Top10%",
    "Anteil Top20%", "Anteil Untere50%", "Alpha", "Schwellenwert Pareto"},
   {avwealth2, totalwealth2, sharetop1, sharetop5, sharetop10, sharetop20,
    sharelower50, finalalpha2, Mean[percvalues2[[All, cutoffpoint2]]]},
   {avwealthc2, totalwealthc2, sharetop1c, sharetop5c,
    sharetop10c, sharetop20c, sharelower50c, "NA", "NA"}};
```

In[78]:=
```
export = Prepend[Transpose[finalresults], {"", "Original", "Schätzung"}] // TableForm
```

Out[78]//TableForm=

	Original	Schätzung
Durchschnitt	258 414.	340 927.
Gesamt	9.9813×10^{11}	1.31684×10^{12}
Anteil Top1%	0.254055	0.405044
Anteil Top5%	0.434437	0.562413
Anteil Top10%	0.555346	0.657897
Anteil Top20%	0.720919	0.785339
Anteil Untere50%	0.0320237	0.0247785
Alpha	1.21773	NA
Schwellenwert Pareto	625 839	NA

```
Export["finalresults.xlsx", export]
```

finalresults.xlsx

Export der geschätzten Daten

```
Export["corrdat2.dat", corrwealth2];
Export["corrweight2.dat", corrweight2];
```