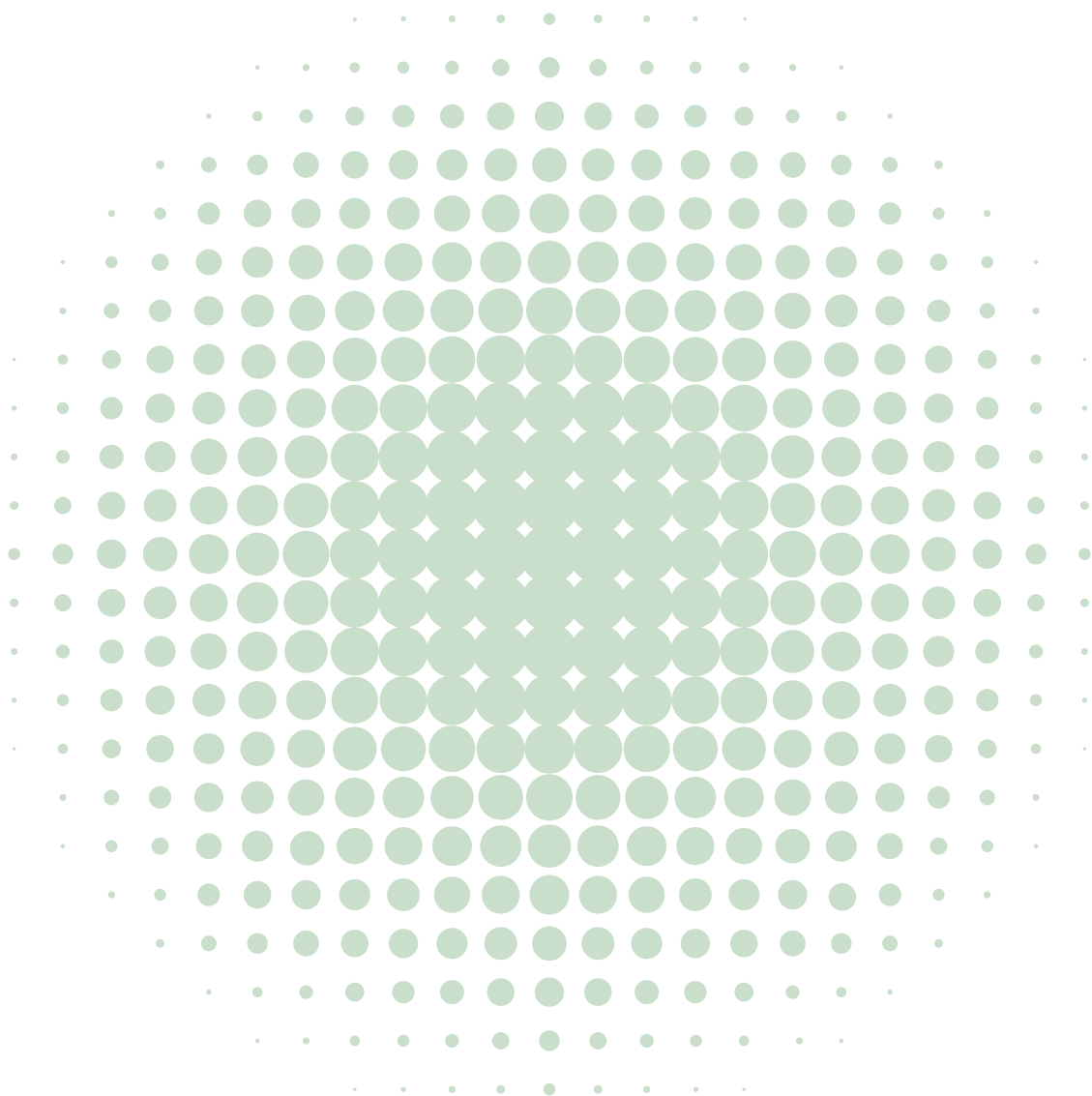




Potenzialabklärung von verknüpften Administrativ- und Befragungsdaten für das nationale Armutsmonitoring

Bericht zuhanden des Nationalen
Armutsmonitorings



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Bundesamt für Sozialversicherungen BSV
Office fédéral des assurances sociales OFAS

Nationales Armutsmonitoring

Das Bundesamt für Sozialversicherungen veröffentlicht konzeptionelle Arbeiten sowie Forschungs- und Evaluationsergebnisse zu aktuellen Themen im Bereich der Sozialen Sicherheit, die damit einem breiteren Publikum zugänglich gemacht und zur Diskussion gestellt werden sollen. Die präsentierten Folgerungen und Empfehlungen geben nicht notwendigerweise die Meinung des Bundesamtes für Sozialversicherungen wieder.

Die vorliegende Publikation wurde im Rahmen des Nationalen Armutsmonitorings erarbeitet. Die gewonnenen Erkenntnisse flossen in den ersten Monitoringbericht 2025 ein.

Autorinnen/Autoren

Dominic Höglinger, Tabea Kaderli, Severin Bischof
Büro für Arbeits- und Sozialpolitische Studien BASS AG
Konsumstrasse 20, CH-3007 Bern
+41 (0)31 380 60 80, info@buerobass.ch
www.buerobass.ch

Auskünfte

Bundesamt für Sozialversicherungen
Effingerstrasse 20, CH-3003 Bern

Geschäftsfeld Familie, Generationen und Gesellschaft
Anna Koukal
+41 (0)58 465 20 78, anna.koukal@bsv.admin.ch

Bereich Forschung und Evaluation
Ilka Steiner
+41 (0)58 483 94 31, ilka.steiner@bsv.admin.ch

Copyright

Bundesamt für Sozialversicherungen, CH-3003 Bern
Auszugsweiser Abdruck – ausser für kommerzielle Nutzung –
unter Quellenangabe und Zustellung eines Belegexemplares
an das Bundesamt für Sozialversicherungen gestattet.

Publikationsdatum und Auflage

November 2025, 1. Auflage

Potenzialabklärung von verknüpften Administrativ- und Befragungsdaten für das nationale Armutsmonitoring

Schlussbericht

Im Auftrag
Bundesamt für Sozialversicherungen BSV

Dominic Höglinger, Tabea Kaderli, Severin Bischof
Bern, 2024 (v1.0)

Inhaltsverzeichnis

Zusammenfassung	II
1 Einleitung	7
2 Diskussion von geeigneten Datenquellen für den verknüpften Datensatz für das Armutsmonitoring	9
2.1 Datenquellen, die bereits im Rahmen von WiSiER verknüpft wurden	9
2.1.1 Mögliche Optimierungen und Ergänzungen	10
2.1.2 Verfügbarer Datenzeitraum für die Analysen	10
2.2 Diskussion möglicher weiterer individueller Datenquellen, die in WiSiER nicht berücksichtigt wurden	11
3 Abdeckung der relevanten Einkommenskomponenten mit dem verknüpften Datensatz	20
3.1 Verfügbare und fehlende Einkommenskomponenten im verknüpften Datensatz	20
3.2 Einkommensinformationen zu spezifischen Personengruppen	22
4 Plausibilisierung der verknüpften Daten ohne Steuerinformationen zur Bestimmung der einkommensschwachen Bevölkerung	23
4.1 Vorgehen bei der Plausibilisierung	23
4.1.1 «Einkommensschwäche» P20 als primär verwendeter Indikator	23
4.1.2 Die verwendeten Varianten der Berechnung des Haushaltseinkommens	25
4.1.3 Grundgesamtheit und Auswahl der relevanten Beobachtungen	26
4.2 Ergebnisse der Plausibilisierung	29
4.2.1 Ergebnisse für die zentrale Gruppierungsvariable Haushaltstyp	31
4.2.2 Ergebnisse zu weiteren Gruppierungsvariablen	38
4.3 Schlussfolgerungen aus den Ergebnissen der Plausibilisierung	44
5 Literaturverzeichnis	47
Anhang	49
A-1 Ergänzende Tabellen und Abbildungen	49

Zusammenfassung

Da zum aktuellen Zeitpunkt kantonale Steuerdaten für eine systematische Verknüpfung nicht zur Verfügung stehen, klärt der vorliegende Bericht das Potenzial der Verknüpfung von Administrativ- und Befragungsdaten *ohne Steuerdaten* für Analysen im Rahmen des nationalen Armutsmonitorings vertieft ab. Im Fokus stehen dabei insbesondere mögliche Datengrundlagen und Auswertungen für den ersten Monitoringbericht, welcher Ende 2025 publiziert wird, und die dort schwerpunktmässig zu behandelnden Themen: Materielle Existenzsicherung, Erwerbsarbeit und Bildung.

Gegenüber stichprobenbasierten Querschnittsdaten (wie die für Analysen zur Armut in der Schweiz zentrale Erhebung über die Einkommen und Lebensbedingungen SILC) liegen die Vorteile von verknüpften Administrativ- und Befragungsdaten darin, dass sie auch **vertiefte Analysen auf subnationaler Ebene, namentlich der Kantone, und Verlaufsanalysen über längere Zeiträume** ermöglichen. Bei Verlaufsanalysen wird die Entwicklung auf Haushalts- oder Personenebene über die Zeit analysiert und es können somit auch Stationen im System der sozialen Sicherheit sowie lebensbiographische, mithin auch intergenerationale Fragestellungen untersucht werden.

Der WiSiER-Datensatz, welcher Steuerdaten von 2012-2015 von 11 Kantonen sowie Administrativ- und Umfragedaten verknüpfte und in zahlreichen Forschungsprojekten zur Analyse der wirtschaftlichen Situation von Personen und Haushalten Verwendung fand, war ein Pionierprojekt, welches das grosse Potenzial von solchen verknüpften Daten aufzeigte (u.a. Wanner 2019, Wanner/Gerber 2021). Der WiSiER-Datensatzes wurde dabei einmalig erstellt und seine Nutzung war zeitlich befristet bis Ende 2023.

Die Verwendung von aktuellen und weitgehend vollständigen kantonalen Steuerdaten, möglichst über einen längeren Datenzeitraum, ist im Rahmen des ersten Berichts des nationalen Armutsmonitorings nicht möglich, auch wenn entsprechende Bemühungen für eine zukünftige Nutzung im Gange sind. In der Potenzialabklärung ist deshalb aufzuzeigen, welche Analysen auch ohne kantonale Steuerdaten mit einem Datensatz aus verknüpften Administrativ- und Befragungsdaten möglich sind und inwiefern die Lücke, welche durch die fehlenden Steuerinformationen besteht, teilweise geschlossen werden kann.

Vor diesem Hintergrund bearbeitet der Bericht die folgende zentrale Frage:

Wie gut eignen sich verknüpfte Administrativ- und Befragungsdaten (ohne kantonale Steuerdaten) zur hinreichend präzisen Identifikation und Quantifizierung der einkommensschwachen Bevölkerung? Diese Frage wird mittels einer umfassenden Plausibilisierung abgeklärt, bei welcher die Ergebnisse basierend auf Administrativ- und Befragungsdaten ohne kantonale Steuerinformationen den Ergebnissen der WiSiER-Referenzdatenbasis (unter Einschluss von Steuerinformationen) gegenübergestellt werden.

Im Rahmen dieser Fragestellung ist vorgängig abzuklären, **welche individuellen Datenquellen sich für einen Einbezug in diesen neuen verknüpften Datensatz anbieten** – insbesondere auch über jene hinaus, welche in WiSiER enthalten waren. Sei dies mit Blick auf Informationen zu relevanten Komponenten des Haushaltseinkommens, als auch hinsichtlich weiterer nützlicher Angaben für Analysevorhaben im Rahmen des Armutsmonitorings.

Geeignete Datenquellen für den verknüpften Datensatz für das Armutsmonitoring

Für den verknüpften Datensatz, mit welchem Analysen im Rahmen des Armutsmonitorings vorgesehen sind, sind die bereits im Rahmen von WiSiER verwendeten folgenden individuellen Datenquellen essentiell:

- Statistik der Bevölkerung und Haushalte STATPOP (zzgl. BEVNAT) des BFS
- Gebäude- und Wohnungstatistik GWS des BFS
- Strukturhebung SE des BFS
- Individuelle Konten AHV-IK von ZAS/BSV (Erwerbseinkommen)
- Rentenregister AHV/IV von ZAS/BSV (insb. Angaben zu Ergänzungsleistungen EL)
- AVAM/ASAL-Daten der Arbeitslosenversicherung des SECO
- Sozialhilfeempfängerstatistik SHS des BFS

Der aufgrund der Verfügbarkeit der individuellen Datenquellen **abgedeckte Datenzeitraum** des verknüpften Datensatzes umfasst die **Jahre ab 2012** bis zum aktuellen Zeitpunkt mindestens 2019, mit verkraftbaren Einbussen bei der Vollständigkeit gewisser Quellen bis 2021. IK-Daten für Angaben zur Erwerbstätigkeit auf individueller Ebene werden für Analysen zusätzlich zurück bis 1982 verfügbar sein.

Verwendet werden soll auch die erst seit kurzem bestehende **STATPOP Haushaltstyp-Variable** des BFS, welche für das Datenjahr 2015 und später verfügbar ist.

Neben diesen bereits im Rahmen von WiSiER bewährten individuellen Datenquellen wurde im Rahmen der Potenzialabklärung auch der Einbezug weiterer geeigneter Datenquellen geprüft. Dabei gilt es auch den analytischen Mehrwert mit dem anfallenden Mehraufwand für die zusätzliche Verknüpfung abzuwägen. Weitere relevante Kriterien, wie ausreichend langer verfügbarer Datenzeitraum, ausreichende Fallzahlen (idealerweise Vollerhebung) und als Grundvoraussetzung die Möglichkeit der Verknüpfung mittels erfasster Sozialversicherungsnummer, sind ebenfalls zu berücksichtigen. Im vorliegenden Bericht in Kapitel 2.2 findet sich dazu eine **systematische tabellarische Übersicht** von für die Verknüpfung geeigneten möglichen Datenquellen.

Namentlich zu erwähnen sind folgende möglichen Erweiterungen, welche für die vorgesehenen Auswertungsvorhaben einen wesentlichen Mehrwert versprechen:

■ **Statistiken zur Sozialhilfe im Asyl- und Flüchtlingsbereich** (SH-FlüStat, SH-AsylStat) des BFS, um die bereits integrierte reguläre Sozialhilfeempfängerstatistik zu komplementieren.

■ **Lohnstrukturerhebung LSE** als alternative Quelle für Angaben zum Erwerbsspensum und zum Lohn.

Weitere individuelle Datenquellen, wie etwa zur Alimentenbevorschussung ALBV des BFS, die EFIND-Daten des BAG mit den obligatorischen Krankenkassenprämien brutto und netto (und somit auch mit Angaben zur Prämienverbilligung) oder das Familienzulagenregister von ZAS/BSV bieten aktuell angesichts der bestehenden grösseren Lücken bei den weiteren Einkommenskomponenten nur einen begrenzten Mehrwert, der sich jedoch bei einem allfälligen zukünftigen Einbezug der kantonalen Steuerdaten markant erhöhen würde.

Verfügbare und fehlende Komponenten zur Bestimmung des Haushaltseinkommens

Für Haushalte mit Personen im erwerbsfähigen Alter ist die primäre Einkommensquelle das **Erwerbseinkommen**. Dieses kann mit dem verknüpften Datensatz befriedigend mittels der Angaben aus den individuellen Konten als Ersatz für die entsprechenden nicht verfügbaren kantonalen Steuerinformationen bestimmt werden.

Aufgrund der nicht verfügbaren Steuerinformationen fehlen im vorgesehen verknüpften Datensatz gegenüber WiSiER Angaben zu

wesentlichen weiteren Einkommenskomponenten, namentlich zu erhaltenen wie auch ausgerichteten **Unterhaltszahlungen** und zu **Renten der beruflichen Vorsorge BVG**. Weiter fehlen auch Angaben zu Vermögen und Vermögensverzehr (wobei zur Berechnung der primären Armutsquote des BFS das Vermögen ebenfalls nicht berücksichtigt wird), zu Krankentaggeldern, sowie zu Kinder- und Ausbildungszulagen (wobei Kinderzulagen bis zum 16. Lebensjahr relativ einfach zu modellieren sind).

In WiSiER nicht enthalten waren Angaben zu einzelnen bedarfsabhängigen Sozialleistungen, wie zur Sozialhilfe im Asyl- und Flüchtlingsbereich, zur Alimentenbevorschussung, zu Ausbildungsstipendien sowie zur Prämienverbilligung. Auch fehlten Angaben zu den obligatorischen Ausgaben, die zur Bestimmung des verfügbaren Haushaltseinkommens wesentlich sind, namentlich zu den Prämien der obligatorischen Krankenversicherung und zu den Wohnkosten.

Wie oben erwähnt liessen sich einzelne dieser bestehenden Lücken mit dem Einbezug zusätzlicher und gegenüber WiSiER neuer Datenquellen schliessen. Eine andere Möglichkeit bestünde allenfalls darin, diese fehlenden Informationen mittels Imputationsverfahren zu vervollständigen.

Herausforderungen mit Blick auf spezifische Bevölkerungssegmente

Ein Vorteil des verknüpften Datensatzes gegenüber WiSiER ist die **umfassende Abbildung der ausländischen Bevölkerung**. Quellbesteuerten Personen, überwiegend Ausländer und Ausländerinnen ohne Niederlassungsbewilligung C, waren aufgrund der fehlenden regulären Steuerveranlagung in den meisten mit WiSiER durchgeführten Analysen nicht enthalten. Weiterhin bestehen bleibt die generell **schwierige adäquate Abbildung der finanziellen Situation von Selbständigen**, und auch **Haushalten von Personen unter 25 Jahren** stellen eine Herausforderung dar. Eine für viele dieser Haushalte wesentliche Einkommensquelle, elterliche Transferzahlungen, ist nicht in den Daten enthalten.

Da in den verknüpften Administrativ- und Befragungsdaten ohne Steuerinformationen keine Angaben zu den BVG-Renten verfügbar sind, welche für einen beträchtlichen Teil der Rentenhaushalte eine substantielle Einkommensquelle darstellen, sind **Haushalte mit Personen im Rentenalter** aus den vorgesehenen Analysen notgedrungen auszuschliessen. Dabei ist auch ein ab 58 Jahren möglicher BVG-Rentenvorbezug zu berücksichtigen. Für die Plausibilisierung wurden deshalb Haushalte mit Personen

über 60 Jahren ausgeschlossen. Durchgeführte empirische Analysen zeigten auf, dass ab Alter 61 des ältesten Haushaltsmitglieds der Anteil an Personen in Haushalten mit einer BVG-Rente 20 Prozent überschreitet (und ohne verfügbare Angaben dazu wesentlich verzerrte Ergebnisse zum Haushaltseinkommen die Folge wären).

Ergebnisse der Plausibilisierung der verknüpften Daten ohne Steuerinformationen

Inwiefern der Wegfall der kantonalen Steuerinformationen (und der dort enthaltenen Einkommenskomponenten) die Fähigkeit zur Identifikation und Quantifizierung der einkommensschwachen Bevölkerung beeinträchtigt, wurde im Rahmen der Potenzialabklärung mittels einer umfassenden Plausibilisierung geprüft. Sämtliche Analysen wurden dabei mit den WiSiER-Daten des jüngsten verfügbaren Jahres 2015 durchgeführt, wobei jeweils die Ergebnisse basierend auf den verknüpften Daten ohne Steuerdaten der Referenzdatenbasis WiSiER mit Steuerdaten gegenübergestellt wurden.

Als **primärer Indikator** wurde dabei die **einkommensschwache Bevölkerung (P20)**, bestimmt als die 20 Prozent der Bevölkerung mit dem tiefsten Haushaltsäquivalenzeinkommen verwendet (vgl. Bio Tchané 2019, Fluder et al. 2020). Alternativ wurde auch für den Grenzwert von 60 Prozent des medianen Haushaltsäquivalenzeinkommens eine Plausibilisierung durchgeführt, der für das relative Armutsmass der Armutsgefährdungsgrenze verwendet wird.

Neben dem Haushaltseinkommen, wie es mit den in den verknüpften Administrativ- und Befragungsdaten ohne Steuerinformationen vorhandenen Komponenten bestimmt werden kann, wurde auch eine **optimierte Variante** (Berechnung des Nettoerwerbseinkommens, Modellierung der Kinderzulagen) sowie eine konzeptionell **nur auf das Erwerbseinkommen fokussierte Variante** (inklusive allfällige Leistungen der Arbeitslosenversicherung) plausibilisiert.

Die Ergebnisse der Plausibilisierung zeigen, dass die verknüpften Datenbasis (ohne Steuerinformationen) gegenüber der Referenzdatenbasis WiSiER mit Steuer einen **Anteil an abweichenden Klassifikationen bezüglich Einkommensschwäche (P20) im Ausmass von rund 10%** aufweist. Konkret bedeutet dies, dass eine von zehn Personen als (nicht) einkommensschwach klassifiziert wird, obwohl sie dies gemäss Referenzdatenbasis WiSiER (nicht) ist. Bei einer sogenannten «fuzzy»-Klassifikation, welche knapp falsch zugeordnete Beobachtungen noch als korrekt behandelt, sind

noch gut 6% der Individuen abweichend klassifiziert. Problematisch ist bei diesen Fehlklassifikationen weniger ihr Ausmass, als vielmehr der Umstand, dass diese Fehlklassifikationen **nicht zufällig, sondern systematisch** erfolgen.

Die systematischen Fehlklassifikationen manifestieren sich als **Verzerrungen, respektive Abweichungen bei den errechneten Anteilswerten** der einkommensschwachen Personen. Die aufgrund der fehlenden Einkommenskomponenten erwarteten stark von systematischen Fehlklassifikationen betroffenen Bevölkerungssegmente bestätigten sich in der Plausibilisierung: **Einelternhaushalte, Haushalte mit Hinterlassenen sowie Haushalte mit IV-Rentenbezug** weisen eine im Vergleich zu anderen Gruppen wesentlich tiefere Klassifikationsgüte und in der Folge grössere Verzerrungen bei den Anteilswerten auf. Konkret kommt es in diesen Fällen zu einer Überschätzung des Anteils einkommensschwacher Personen. Hauptursache sind die aufgrund der nicht verfügbaren Steuerdaten fehlenden Informationen zu den für diese Bevölkerungssegmente besonders relevanten Einkommenskomponenten **BVG-Renteneinkünfte und Unterhaltszahlungen**.

Inwiefern solche fehlenden Einkommensbestandteile mittels geeigneter Annahmen oder Imputationsverfahren vervollständigt werden können, wurde im Rahmen des Mandats nicht vertieft untersucht. Grundsätzlich stossen solchen Verfahren für den befriedigenden Ersatz von vollständig und nicht nur in Einzelfällen fehlenden Angaben, die auch noch stark individuell variabel sind, schnell an Grenzen.

Mit der Plausibilisierung konnten die Verzerrungen quantifiziert und somit ihr genaues Ausmass beziffert werden. Eine Verwendung dieses Indikators zur Einkommensschwäche ohne Weiteres basierend auf den verknüpften Daten ohne Steuer für Analysen im Rahmen des Armutsmonitorings scheint vor diesem Hintergrund fraglich.

Ein **Ausschluss** der aus methodischer Sicht problematischen Bevölkerungssegmente, wie dies als mögliche Lösung gesehen werden könnte, ist **nicht zielführend**. Dies da einerseits bekannte Risikogruppen, wie Alleinerziehende und IV-Rentenbeziehende, aus den Analysen ausgeschlossen würden, die für die Armutsdebatte besonders relevant sind. Zum anderen hätte ein Ausschluss bestimmter Gruppen schwer kontrollierbare Folgen auch auf weitere Analysen und deren Validität, da andere Kategorisierungen und deren Kategorien je nach ihrer Zusammensetzung mit Blick auf diese ausgeschlossenen Gruppen unterschiedlich betroffen, respektive verzerrt würden. Insbesondere Entwicklungen von Anteilswerten im

zeitlichen Verlauf sind aufgrund von Kompositionseffekten anfällig auf solche Probleme. Auch **bereits geringfügige Abweichungen sind problematisch**, da bei Kennzahlen zur Armut wie der Armutsquote auch geringfügige Veränderungen über die Zeit oder zwischen Bevölkerungsgruppen von auch nur einem oder wenigen Prozentpunkten substantiell relevant sind.

Dies heisst nicht, dass Auswertungen grundsätzlich nicht sinnvoll sind. Hingegen sind die entsprechenden Herausforderungen jeweils bei der Analyse zu reflektieren. Auch sind die Limitationen bei der Vermittlung der Befunde transparent auszuweisen.

Schlussfazit

Insgesamt zeigt die vorgenommene Plausibilisierung, dass die Nichtverfügbarkeit von kantonalen Steuerinformationen die Verwendung der verknüpften Administrativ- und Befragungsdaten zur Analyse von Armut in der Schweiz **nicht verunmöglicht, deren Nutzen für das Armutsmonitoring jedoch wesentlich schmälert**.

Ein Auftrag an das nationale Armutsmonitoring ist die Durchführung kantonal vergleichender Analysen. Um Kennzahlen wie kantonale Armutsquoten vergleichend darzustellen, sind Steuerdaten notwendig. Die **Verfügbarkeit von kantonalen Steuerdaten** bleibt vor diesem Hintergrund ein **zentrales und dringendes Erfordernis für valide Analysen zur Armut auf kantonaler Ebene**. Die Steuerdaten sind dabei im Idealfall auch für zurückliegende Jahre verfügbar, um Analysen von längeren Zeitreihen zu ermöglichen.

Liegt hingegen der Fokus einer Fragestellung nicht auf Armut oder die Einkommenssituation gesamthaft, sondern interessiert spezifisch Erwerbseinkommensschwäche oder etwa der Bezug von Sozialhilfeleistungen (namentlich Sozialhilfe), insbesondere auch in der zeitlichen Entwicklung, so bieten die verknüpften Administrativ- und Befragungsdaten auch ohne Steuerdaten ein beträchtliches Potenzial für aufschlussreiche und valide Analysen. Entsprechende Analysevorhaben werden in einem Folgeman dat näher ausgearbeitet und zur Verwendung im Rahmen des Armutsmonitorings anschliessend durchgeführt.

1 Einleitung

Das Parlament hat den Bundesrat im Juni 2020 mit der Überweisung der Motion 19.3953 der ständerrätlichen Kommission für Wissenschaft, Bildung und Kultur beauftragt, ein nationales Armutsmonitoring aufzubauen. Ein erster Bericht ist für Ende 2025 geplant. Das Monitoring soll die Armutslage der Bevölkerung detailliert beschreiben und zugleich aufzeigen, was über die Wirksamkeit von Strategien zur Bekämpfung und Prävention von Armut bekannt ist.

Im Rahmen des Armutsmonitorings wird die Armutssituation der Schweizer Bevölkerung anhand von statistischen Indikatoren beschrieben. Verknüpfte Administrativ- und Befragungsdaten ermöglichen es dabei gegenüber den stichprobenbasierten Querschnittsdaten, wie insbesondere die für Analysen zur Armut in der Schweiz zentrale Erhebung über die Einkommen und Lebensbedingungen SILC des Bundesamts für Statistik, weitere essentielle Aspekte auszuleuchten. Die Vorteile von verknüpften Administrativ- und Befragungsdaten liegen insbesondere darin, dass sie **vertiefte Analysen auf subnationaler Ebene, namentlich der Kantone**, sowie **Verlaufsanalysen über längere Zeiträume** ermöglichen. Bei Verlaufsanalysen wird die Entwicklung auf Haushalts- oder Personenebene über die Zeit analysiert und es können somit auch Stationen im System der sozialen Sicherheit sowie lebensbiographische, mithin auch intergenerationale Fragestellungen untersucht werden. Auch ermöglichen diese Daten die Analyse von interessierenden Bevölkerungsgruppen, welche in den stichprobenbasierten Daten nicht in genügend hoher Fallzahl enthalten sind.

Der WiSiER-Datensatz, welcher Steuerdaten von 2012-2015 von 11 Kantonen sowie Administrativ- und Umfragedaten verknüpfte und in zahlreichen Forschungsprojekten zur Analyse der wirtschaftlichen Situation von Personen und Haushalten Verwendung fand, war ein Pionierprojekt, welches das grosse Potenzial von solchen verknüpften Daten aufzeigte (u.a. Wanner 2019, Wanner/Gerber 2021). Der WiSiER-Datensatzes wurde dabei einmalig erstellt und seine Nutzung war zeitlich befristet bis Ende 2023.

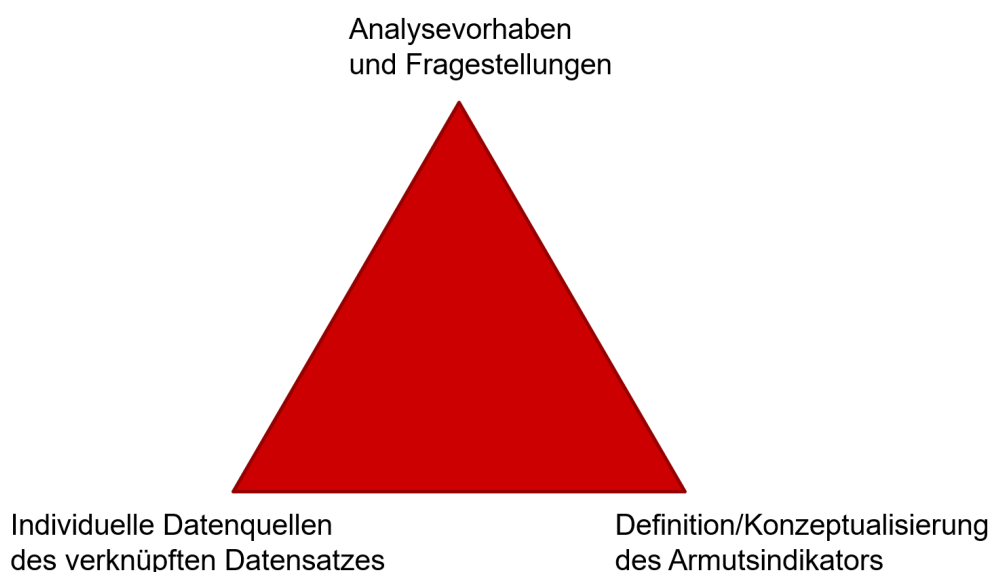
Der Einbezug kantonaler Steuerinformationen ist für Analysen basierend auf verknüpften Daten zu Armut und zur Einkommenssituation essentiell und deshalb auch für das nationale Armutsmonitoring langfristig angedacht. Es sind Bemühungen im Gange, solche kantonalen Steuerdaten für zukünftige Analysen im Rahmen des Armutsmonitorings nutzen zu können. Im Moment stehen diese jedoch noch nicht zur Verfügung. Und auch wenn diese Bemühungen in absehbarer Zukunft von Erfolg gekrönt sein werden, so werden kantonale Steuerdaten über mehrere Jahre hinweg, wie sie zur Durchführung von Längsschnittanalysen benötigt werden, noch längere Zeit nicht verfügbar sein. Es sind deshalb alternative Vorgehensweisen ohne Verwendung von kantonalen Steuerdaten zu prüfen.

Vor diesem Hintergrund verfolgt das diesem Bericht zugrundeliegende Mandat zur Potentialabklärung von verknüpften Administrativ- und Befragungsdaten für das nationale Armutsmonitoring die folgenden Fragestellungen:

- (1.) Das **Potential von verknüpften Administrativ- und Befragungsdaten** (ohne die bis auf weiteres nicht verfügbaren kantonalen Steuerdaten) **zur hinreichend präzisen Identifikation, Beschreibung und Quantifizierung der einkommensschwachen Bevölkerung** als Voraussetzung für die oben erwähnten Analysevorhaben ist vertieft abzuklären.
- (2.) Dabei ist in einem vorgängigen Schritt zu prüfen, **welche individuellen Datenquellen sich für einen Einbezug in diesen neuen verknüpften Datensatz anbieten** – insbesondere auch über jene hinaus, welche im Rahmen des WiSiER-Projektes bereits enthalten waren. Sei dies mit Blick auf Informationen zu relevanten Komponenten zur Bestimmung des Haushaltseinkommens, als auch hinsichtlich weiterer nützlicher Angaben für Analysevorhaben im Rahmen des Armutsmonitorings.

Der vorliegende Bericht mit den Ergebnissen dieser Potentialabklärung liefert die Grundlagen, um diese Fragen fundiert beantworten zu können. In einem Folgemandat sollen ausgehend von diesen Ergebnissen konkrete Auswertungsvorhaben mit diesen verknüpften Daten näher ausgearbeitet und zur Verwendung im Rahmen des Armutsmonitorings anschliessend durchgeführt werden. Dabei können die drei bearbeiteten Bereiche des Mandats – Prüfung der Eignung von individuellen Datenquellen für die Integration in den verknüpften Datensatz; Bestimmung des zu verwendbaren Indikators (inkl. Definition und Konzeptualisierung); Ausarbeitung von konkreten Analysevorhaben im Rahmen des Folgemandats – nicht losgelöst von einander betrachtet werden, sondern stehen in enger Wechselwirkung zueinander, wie dies **Abbildung 1** illustriert.

Abbildung 1: Schaubild Zusammenhang der drei Bereiche des Projektmandats



Darstellung BASS

Der Bericht ist wie folgt aufgebaut: **Kapitel 2** bietet eine systematische Übersicht und Diskussion der verfügbaren individuellen Datenquellen, die in den zu erstellenden verknüpften Datensatz integriert werden können, inklusive des sich jeweils ergebenden Mehrwerts, sowie der zu beachtenden Herausforderungen und Einschränkungen. In **Kapitel 3** wird dargelegt, welche relevanten Komponenten des Haushaltseinkommens im zu erstellenden verknüpften Datensatz verfügbar sind und welche Lücken bestehen. Auch wird auf spezifische Herausforderung zur Bestimmung des Haushaltseinkommens bei bestimmten Personengruppen, wie etwa quellenbesteuerte Ausländerinnen und Ausländer oder AHV-Selbständige eingegangen. Auch die für die Analysen aufgrund der Datenlage verfügbaren Datenzeiträume werden diskutiert. **Kapitel 4** schliesslich enthält das zentrale Element des Berichts, die Ergebnisse einer umfassenden Plausibilisierung der neuen verknüpften Datenbasis (ohne Steuerdaten) zur Identifikation und Quantifizierung der «einkommensschwachen Bevölkerung», wobei verschiedene Varianten zur Bestimmung des Haushaltseinkommens plausibilisiert werden – primär im Vordergrund steht der Indikator «Einkommensschwäche» P20, als Datengrundlage wird zusätzlich auch eine optimierte Variante sowie eine Variante mit Fokus nur auf das Erwerbseinkommen plausibilisiert. Das Kapitel schliesst mit einem Abschnitt mit Schlussfolgerungen, welche sich aus diesen Ergebnissen ableiten. Während dieses Kapitel eine Auswahl der wichtigsten Ergebnisse präsentiert und diskutiert, finden sich sämtliche Ergebnisse der Plausibilisierung in einem **separaten Auswertungsbericht**.

2 Diskussion von geeigneten Datenquellen für den verknüpften Datensatz für das Armutsmonitoring

Im Folgenden werden für die Verknüpfung und anschliessende Analyse im Rahmen des nationalen Armutsmonitorings potentiell geeignete individuelle Administrativ- und Befragungs-Datenquellen vorgestellt. Ausgangspunkt dazu bilden die in WiSiER bereits integrierten individuellen Datenquellen - dies ohne die in WiSiER verfügbaren harmonisierten kantonalen Steuerdaten, welche bis auf weiteres nicht mehr zur Verfügung stehen. Die bereits in WiSiER verwendeten Datenquellen werden in Abschnitt 2.1 aufgeführt und kurz diskutiert. In diesem Abschnitt werden auch Optimierungen und Ergänzungen vorgeschlagen, die aufgrund von Neuerungen und aufgrund des erzwungenen Verzichts auf die Verwendung von kantonalen Steuerdaten möglich werden. Auch die für die Analyse verfügbaren unterschiedlichen Zeiträume («Datenjahre») werden dargelegt.

In Abschnitt 2.2 werden dann potentiell geeignete neue Administrativ- und Befragungs-Datenquellen aufgeführt und ihre Eignung diskutiert. Dies sowohl mit Blick auf geeignete Analysen für den ersten Monitoringbericht als auch längerfristig, wenn allenfalls auch die kantonalen Steuerdaten wieder verfügbar sind. Der Abschnitt enthält auch eine systematische Übersicht über alle potentiellen individuellen Datenquellen in Form einer Tabelle.

2.1 Datenquellen, die bereits im Rahmen von WiSiER verknüpft wurden

WiSiER basiert, neben den harmonisierten kantonalen Steuerdaten, auf folgenden weiteren Administrativ- oder Befragungsdaten:

■ **Statistik der Bevölkerung und der Haushalte STATPOP** des BFS (Bestände und Bewegungen), zzgl. der «verwandten» Datensätze **Gebäude- und Wohnungsstatistik GWS** und **Statistik der natürlichen Bevölkerungsbewegung BEVNAT**.

■ **Strukturerhebung SE** des BFS. Es handelt sich dabei um eine Stichprobe, mittels Pooling über mehrere Jahre lassen sich aber für eine grosse Anzahl an Individuen bestimmte interessierende zusätzliche Informationen bestimmen (namentlich Bildungsstand, berufliche Branche, wöchentliche Arbeitszeit). Die Variablen werden für die Zielperson, für den Haushalt, und tlw. auch für die weiteren Haushaltsmitglieder erfasst.

■ **Daten der individuellen Konten AHV-IK** der ZAS, welche die Abbildung der langfristigen Erwerbsbiographie und der Entwicklung des Erwerbseinkommens ermöglicht (ab 1982). Nicht verfügbar sind dabei Angaben zum Erwerbspensum.

■ **Rentenregister AHV/IV sowie Ergänzungsleistungsregister** von ZAS/BSV (inkl. **Hilflosenentschädigung HE**). Zu beachten ist, dass allfällige Alters- und Invaliditäts-Rentenleistungen der beruflichen Vorsorge (BVG) nicht enthalten sind.

■ **Daten zu Leistungen der Arbeitslosenversicherung AVAM/ASAL** des SECO.

■ **Schweizerische Sozialhilfeempfängerstatistik SHS** des BFS.

Diese Datenquellen haben sich im Rahmen von WiSiER bewährt und sollen deshalb auch Teil des neu zu erstellenden verknüpften Datensatzes sein. Grundsätzlich ist es sinnvoll, sich dabei an den für WiSiER aus diesen Datensätzen verwendeten Variablen und Teildatensätzen zu orientieren (vgl. Bestellformular WiSiER «codebook-pool.xlsx»). Bei Bedarf sind punktuelle Ergänzungen um weitere Variablen möglich. Konkret etwa die neue imputierte Variable Haushaltstyp in STATPOP (vgl. unten).

2.1.1 Mögliche Optimierungen und Ergänzungen

Verwendung der neuen Haushaltstypologie in STATPOP

Das BFS erstellt neu nicht nur für die Strukturerhebung, sondern auch für STATPOP eine Variable «Haushaltstyp». Diese Variable in STATPOP wird imputiert und im Moment noch als experimentelle Statistik geführt, wobei die fehlerhaften Zuweisungen weniger als 1,5% aller Haushalte betreffen und weitere Optimierungen in Arbeit sind.¹ Für WiSiER-Analysen wurde bisher eine eigens erstellte Haushaltsvariable verwendet, grundsätzlich aber nach den gleichen Prinzipien – es sind keine substantiellen Unterschiede zwischen den beiden Herangehensweisen zu erwarten. Im Sinne der Einheitlichkeit der Analysen ist anzustreben, diese neue STATPOP Haushaltstyp-Variable des BFS zu verwenden. Diese ist jedoch erst ab dem Jahr 2015 verfügbar. Im Rahmen der Analysen ist deshalb zu bestimmen, wie die Haushaltsbildung für die früheren drei Jahre 2012-2014 des angestrebten Datenzeitraums vorzunehmen ist.

Kalenderjährliche Betrachtungsweise nicht mehr zwingend

Die aufgrund der Steuerdaten bei WiSiER-Analysen eingenommene (kalender-)jährliche Betrachtungsweise ist mit dem Wegfall dieser Datenquelle nicht mehr zwingend. Obwohl eine kalenderjährliche Betrachtungsweise bei den meisten Analysen grundsätzlich weiterhin sinnvoll sein dürfte (sowohl inhaltlich, als auch bezüglich Aufwand/Ertrag und Datensatz-Handling), so kann bei Bedarf auch eine feinere zeitliche Gliederung verwendet werden. Die meisten anderen Datenquellen zeichnen auch unterjährige Entwicklungen präzise auf, meist in Form von einzelnen Monaten. Phänomene wie unregelmässige Erwerbstätigkeit, unterjährige Ausfälle des Erwerbseinkommens (prekäre Erwerbstätigkeit), Sozialleistungsbezug (namentlich Sozialhilfe) usw. können so bei Bedarf besser/präziser berücksichtigt und prekäre Situationen besser identifiziert werden. Generell wird eine feinkörnigere Analyse von Verläufen möglich.

2.1.2 Verfügbarer Datenzeitraum für die Analysen

Eine Haushalts-ID in für das BFS befriedigender Qualität ist in STATPOP ab 2012 verfügbar. Ab dem Jahr 2012 ist deshalb eine verlässliche Haushaltsbildung möglich. Entsprechend wurden auch mit WiSiER die meisten Zeitreihenanalysen ab 2012 durchgeführt.² Auch für den verknüpften Datensatz sind deshalb **primär Analysen ab 2012 bis zum jüngsten verfügbaren Datenjahr** zum Zeitpunkt des Abzugs durchführbar. Eine Ausnahme bilden die IK-Daten. Diese sind bereits ab 1982 verfügbar und erlauben für einen wesentlich längeren Zeitraum Analysen zur Erwerbsbiografie (Erwerbstätigkeit, Erwerbseinkommen und Erwerbsunterbrüchen) – allerdings nur auf individueller Personenebene und nicht auf der Haushaltsebene.

Während für den WiSiER-Datensatz aufgrund der Verfügbarkeit der Steuerdaten eine relativ kurze vollständige Zeitreihe zur Verfügung stand, von 2012 bis 2015, so ist diese im verknüpften Datensatz deutlich länger. Das jüngste verfügbare Datenjahr ergibt sich jeweils nach Verfügbarkeit der individuellen Datenquelle zum Zeitpunkt der Datenaufbereitung/Verknüpfung – zwischen 1-3 Jahre nach dem Erhebungsjahr (bei den IK-Daten sind zudem stark verzögerte Eintragungen bei Selbständigen zu beachten). Ein Abzug der IK-Daten im März 2024 würde basierend auf den Angaben gemäss BSV (2023b) weitgehend vollständige Angaben der Selbständigen (mit wenigen Lücken) des Jahres 2019 umfassen. Im Rahmen der kantonalen Armutsberichterstattung für den Kanton Luzern wurde von der kantonalen Statistikbehörde LUSTAT auch der Abzug von IK-Angaben zu späteren Jahren

¹ Vgl. die entsprechende Webpage des BFS: <https://www.experimental.bfs.admin.ch/expstat/de/home/projekte/hhtyp.html>

² Dies auch wenn die individuellen Datensätze im WiSiER-Pool meist bereits ab 2010 verfügbar sind.

(bis 3 Jahre vor Abzug) als qualitativ befriedigend erachtet. Abgesehen von den IK-Daten sind die Daten der Schweizerischen Sozialhilfestatistik vermutlich der «Flaschenhals», sie sind jeweils ab Frühjahr des übernächsten Jahres nach dem Erhebungsjahr verfügbar.

Je nach genauem Zeitpunkt und Dauer der Aufbereitung/Verknüpfung sowie Anforderungen an die Vollständigkeit der Daten dürften für den ersten Monitoringbericht somit die **Jahre ab 2012 bis mindestens 2019, mit gewissen Einbussen bei der Vollständigkeit bis 2020** (COVID-1 Pandemiebeginn Schweiz, bedingt geeignet als Referenzjahr für Querschnittsanalysen) **oder gar 2021 für Analysen verwendbar** sein. Für Analysevorhaben zum lebensbiographischen Verlauf und insbesondere zur Transition ins Erwachsenenesein ist eine Zeitreihe, welche diese Übergangsphase möglichst vollständig abdeckt, essenziell und somit jedes zusätzliche Jahr wertvoll.³

2.2 Diskussion möglicher weiterer individueller Datenquellen, die in WiSiER nicht berücksichtigt wurden

Neben den in WiSiER bereits integrierten Datenquellen ist auch der Einbezug weiterer potentiell geeigneter Datenquellen zu prüfen. Für den Entscheid, ob und welche weiteren individuellen Datensätze ergänzend einzubeziehen sind, ist das wichtigste Kriterium der **Mehrwert bezüglich ergänzender Analysen zu relevanten Fragestellungen im Rahmen des Armutsmonitorings**, die so nicht bereits mittels der individuellen Datenquellen alleine (unverknüpft) oder mittels alternativer, bereits verfügbarer Daten (bspw. SILC) beantwortet werden können. Zu berücksichtigen ist auch der anfallende Aufwand für den Einbezug der weiteren Datenquelle (u.a. Beschaffung und Aufbereitung, steigende Komplexität des Datensatzes mit steigender Anzahl Datenquellen).

Daneben gilt es bei der Beurteilung des Potenzials einer individuellen Datenquelle folgende weiteren Kriterien zu berücksichtigen:

- **Verfügbarer ausreichend langer Datenzeitraum:** Ist die potentielle Datenquelle über den gesamten Datenzeitraum des vorgesehenen verknüpften Datensatzes (ab 2012) verfügbar? Da die Analyse von zeitlichen Verläufen ein Schwerpunkt der vorgesehenen Analysen ist, reduziert die Verfügbarkeit von nur jüngsten Datenjahren bei einer potentiellen Datenquelle den Nutzen erheblich.
- **Vollerhebung oder mindestens grosse Stichprobe:** Die vorgesehenen differenzierten Analysen (u.a. nach Kantonen, kleinen Bevölkerungsgruppen) bedingen eine Vollerhebung oder mindestens eine grosse Stichprobe. Verlaufsanalysen sind – ausser bei Stichproben-Paneldaten – zudem grundsätzlich nur bei einer Vollerhebung möglich. Im WiSiER-Datensatz wird aktuell als einzige Stichprobe die Strukturerhebung einbezogen, wobei faktisch dort nur die zeitlich nicht variable und über mehrere Jahre gepoolte Information zum höchsten Bildungsabschluss für Analysen verwendet wurde.
- Eine grundsätzliche Voraussetzung für den Einbezug weiterer Datenquellen ist deren potentielle Verknüpfbarkeit, welche die dazu notwendige **Verfügbarkeit der Sozialversicherungsnummer** bedingt. Die SV-Nr. ist in der Regel bei staatlichen Registern und Befragungen des Bundesamts für Statistik BFS vorhanden. Bei anderen Datenquellen, etwa bei nicht vom BFS verantworteten wissenschaftlichen Befragungen, steht sie nicht zur Verfügung. Damit engt sich das mögliche Universum an verfügbaren zu verknüpfenden Datenquellen bereits zum Vornherein beträchtlich ein. Hinzu kommen

³ Diese Transitionsphase von Kindheit bis zum Erwachsenenesein erstreckt sich gemäss der einschlägigen Literatur vom 14. bis zum 30. Lebensjahr und umfasst u.a. Beginn und Abschluss einer nachobligatorischen Ausbildung, Einstieg ins Erwerbsleben und stabile Erwerbstätigkeit, Partnerschaft und Familiengründung. Die verfügbare Zeitreihe ist damit im Moment noch zu kurz, um diese lebensbiographische Phase für eine einzelne Kohorte vollständig abzudecken.

allfällige **datenschutzrechtliche Einschränkungen**, auch vor dem Hintergrund des neuen Datenschutzgesetzes.

Die Daten der **Erhebung über die Einkommen und Lebensbedingungen (SILC)** decken die zahlreichen Dimensionen von Armut bereits umfassend ab und bilden den Kern der für das Armutsmonitoring vorgesehenen Analysen. Die verknüpften Administrativ- und Befragungsdaten sollen ergänzende Analysen (insbesondere Verlaufsanalysen und Analysen für spezifische Zielgruppen sowie kantonale Vergleiche) dazu ermöglichen, weshalb auf diesen Datensatz und dessen Möglichkeiten im Folgenden nicht näher eingegangen wird.

Tabelle 1 liefert eine systematische Übersicht über die grundsätzliche Eignung, die thematische Abdeckung sowie das Potenzial von allfälligen neu einzubeziehenden individuellen Datenquellen. Der Vollständigkeit halber sind zu Beginn auch die bereits in WiSiER enthaltenen Datenquellen aufgeführt. Rot hervorgehobene Zellen beinhalten relevante Limitationen bei einzelnen Datensätzen.

2 Diskussion von geeigneten Datenquellen für den verknüpften Datensatz für das Armutsmonitoring

Tabelle 1: Bestehende und potentielle zusätzliche Datenquellen für den Einbezug in den verknüpften Administrativ- und Befragungsdatensatz für das Armutsmonitoring

Datenquelle	Grundsätzliche Eignung /Grundmerkmale			Thematische Bereiche		Bemerkungen	Einschätzung Potenzial für Analysen
	Vollerhebung /Sample	Periodizität	Zeitraum	Einkommenskomponente	Armutsdimension		
Bereits in WiSiER vorhanden:							
Statistik der Bevölkerung & Haushalte STATPOP (zzgl. BEVNAT) BFS	Vollerhebung	kontinuierlich	ab 2012 oder früher	-	-	Basis-Datenquelle Neue Haushaltstypologie des BFS zu übernehmen	Hoch, beizubehalten
Gebäude- und Wohnungsstatistik (GWS BFS)	VE	kontinuierlich	ab 2012 oder früher	-	Wohnen	Basale Analysen der (formalen) Wohnverhältnisse (u.a. Belegungsdichte, Fläche pro Person, Wohnungsalter und -typ) sind mit der GWS (sowie SE) möglich, in SILC stehen jedoch gehaltvollere Indikatoren zur Wohnqualität zur Verfügung (u.a. Lärm, Feuchtigkeit, Vandalismus).	Mittel, beizubehalten
Strukturerhebung (SE BFS)	Sample gross (200'000), Querschnitt	kontinuierlich	ab 2012 oder früher	-	Bildung, Wohnen, Erwerbstätigkeit	Höchste abgeschlossene Ausbildung (von Zielperson sowie weiteren HH-Mitgliedern als zentrale Information). Nutzung von Angaben zu wöchentlichen Arbeitsstunden (von Zielperson sowie weiteren HH-Mitgliedern) Generelle Einschränkung: Stichprobe, Längsschnittanalysen bei zeitvarianten Variablen (u.a. Arbeitsstunden) limitiert	Hoch, beizubehalten
Daten der individuellen Konten AHV-IK (BSV/ZAS)	Vollerhebung	kontinuierlich	ab 1982	Brutto-Erwerbseinkommen (inkl. ALV)	Erwerbstätigkeit	Keine Angaben zum Erwerbsspensum. AHV-Selbständige werden oft stark verspätet erfasst.	Hoch, beizubehalten
Rentenregister AHV/IV (BSV/ ZAS)	VE	kontinuierlich	ab 2012 oder früher	Sozialleistung (AHV, IV, EL, HE)	Existenzsicherung	-	Hoch, beizubehalten
Daten Arbeitslosenversicherung (AVAM/ASAL SECO)	VE	kontinuierlich	ab 2012 oder früher	Sozialleistung	Erwerbstätigkeit	-	Hoch, beizubehalten
Sozialhilfeempfängerstatistik (SHS BFS)	VE	kontinuierlich	ab 2012 oder früher	Sozialleistung	Existenzsicherung	-	Hoch, beizubehalten
Potenzielle neue Datenquellen:							
Sozialhilfe Flüchtlingsbereich SH-FlüStat (SHS BFS)	VE	kontinuierlich	ab 2012 oder früher	Sozialleistung	Existenzsicherung	Schliesst Lücke bei der Sozialhilfe im Asyl- und Flüchtlingsbereich	hoch
Sozialhilfe Asylbereich SH-AsylStat (SHS BFS)	VE	kontinuierlich	Erst ab 2016	Sozialleistung	Existenzsicherung	Schliesst Lücke bei der Sozialhilfe im Asyl- und Flüchtlingsbereich	hoch
Alimentenbevorschussung ALBV (SHS BFS)	VE, ohne Jura	kontinuierlich	ab 2012 oder früher	Sozialleistung	Existenzsicherung	Relevante Mindestsicherungsleistung v.a. für armutsgefährdete Eineltern-Haushalte. Angaben zu normalen Alimentenzahlungen fehlen jedoch weiterhin.	mittel (abhängig von Fragestellung)

2 Diskussion von geeigneten Datenquellen für den verknüpften Datensatz für das Armutsmonitoring

Familienzulagenregister (ZAS)	VE	kontinuierlich	ab 2012	Sozialleistung (Kinder- und Ausbildungszulagen)	-	Informationen zu entrichteten Ausbildungszulagen fehlen bisher (Kinderzulagen können modelliert werden) Nur Bezug und zuständige Kasse, nicht Höhe der Zulage wird erfasst (modellierbar).	Kurzfristig mittel, langfristig hoch (in Kombination mit Steuerdaten)
EFIND-Daten BAG (oblig. KK-Prämien brutto und netto)	VE	kontinuierlich	Erst ab 2019	Sozialleistung (Prämienverbilligung) Ausgabe für verfügbares Einkommen (oblig. Prämien)	-	Einbezug der (gebundenen) Prämienverbilligung sollte nur parallel mit Berücksichtigung der Ausgaben für die oblig. KK-Prämien geschehen. Im Moment nur kurze Zeitreihe verfügbar.	Kurzfristig mittel, langfristig hoch (in Kombination mit Steuerdaten, bei längerer Zeitreihe)
Stipendienregister (STIP Statistik der kantonalen Stipendien und Darlehen BFS)	VE	kontinuierlich	ab 2012 eingeschränkt, ab ~2015 weitgehend vollständig	Sozialleistung	Bildung, Existenzsicherung	Daten können gewisse Lücken bezüglich Einkommen schliessen, u.a. bei armutsbetroffenen Haushalten, in denen die Kinder Stipendien beziehen. Grösseres Problem der nicht erfassten privaten Transferzahlungen durch die Eltern bei Studierenden, die ausserhalb des Elternhaushaltes wohnen, bleibt jedoch bestehen.	mittel
Schweizerische Gesundheitsbefragung (SGB BFS)	Sample klein, Querschnitt	5-jährlich	ab 2012 oder früher	-	Gesundheit, soziale Beziehungen	Zahlreiche Items zu Gesundheit, sozialen Beziehungen sowie SES-Variablen (Einkommen und Erwerb, Bildung); aber Stichprobe Mehrwert der Verknüpfung gegenüber der eigenständigen Nutzung der SGB für das Armutsmonitoring unklar.	gering
Schweizerische Arbeitskräfteerhebung (SAKE BFS, inkl. unregelmässige Spezialmodule wie Vereinbarkeit, Migration, unbezahlte Arbeit)	Sample gross (> 100'000)	kontinuierlich	-	-	Erwerbstätigkeit	Fokus Erwerbstätigkeit sowie Spezialthemen; aber Stichprobe Mehrwert der Verknüpfung gegenüber der eigenständigen Nutzung der SAKE im Rahmen des Armutsmonitoring unklar.	gering
Lohnstrukturhebung (LSE BFS)	Sample gross (2.1 Mio, ohne kleine Betriebe)	2-jährlich	ab 2012	-	Erwerbstätigkeit	Potentiell geeignete und hoch relevante Variablen zum (vertraglichen) Erwerbspensum. Diverse Herausforderungen (Umgang mit unvollständig erfassten Mehrfachanstellungen, variablen/atypischen Arbeitsverhältnissen, fehlenden Selbständigen usw.)	gering, v.a. falls Arbeitspensum alternativ aus SE übernommen wird
Längsschnittanalysen im Bildungsbereich (LABB BFS)	Vollerhebung (mehrheitlich), verknüpfter Datensatz	kontinuierlich	je nach Teildatenquelle ab 2012, oft später	-	Bildung, Erwerbstätigkeit	Der Datensatz umfasst neben den bildungsspezifischen Daten bereits SE, STATPOP, IK und ALV-Informationen. Damit sind zahlreiche relevante Analysen möglich, der Mehrwert einer Verknüpfung scheint gering. Diese wäre allenfalls punktuell und mit Blick auf konkrete spezifische Fragestellungen sinnvoll.	gering
Mikrozensus Aus- und Weiterbildung (xx)	Sample klein, Querschnitt	5-jährlich	2011, 2016, 2021	-	Bildung	Mehrwert der Verknüpfung gegenüber der eigenständigen Nutzung bei Bedarf im Armutsmonitoring unklar. Diese wäre allenfalls punktuell und mit Blick auf konkrete spezifische Fragestellungen sinnvoll.	gering

2 Diskussion von geeigneten Datenquellen für den verknüpften Datensatz für das Armutsmonitoring

Diverse Datenquellen politische Partizipation	-	-	-	-	Politische Partizipation	Administrative Daten zur Partizipation auf kommunaler Ebene wurden in einzelnen wissenschaftlichen Projekten aufbereitet, national sind solche Daten nicht vorhanden. SILC verfügt über Items zur politischen Partizipation (Interesse, Institutionenvertrauen, Partizipation), die u.a. nach Einkommensgruppen ausgewertet werden können. Ansonsten stehen die üblichen politikwissenschaftlichen Datensätze (Selects, ESS, ISSP/MOSAICH) zur Verfügung, die jedoch nicht verknüpfbar sind.	-	-
Harmonisierte kantonale Steuerdaten	Vollerhebung	kontinuierlich	offen		BVG-Renten, Alimente für minderjährige Kinder, weitere Einkommensquellen; Vermögen (Modellierung Vermögensverzehr); Ausgabe für verfügbares Einkommen (Modellierung Steuern)	Bis auf weiteres sind die kantonalen Steuerdaten nicht verfügbar, auch wenn entsprechende Bemühungen laufen, diese längerfristig einzubeziehen. Die Steuerdaten bieten einen unschätzbaren Mehrwert, der von keiner anderen Datenquelle erreicht wird, da erst mit ihnen die umfassende Erfassung aller Einkommenskomponenten möglich ist. Auch Angaben zu Vermögen sind damit verfügbar.	Hoch	

Darstellung BASS.

Im Folgenden vertiefende Ausführungen zu einzelnen in der Tabelle erwähnten Datensätzen:

■ Im Rahmen der regulären Sozialhilfestatistik werden sozialhilfebeziehende Flüchtlinge unter 5 Jahren Aufenthalt sowie vorläufig aufgenommene Personen (VAP) unter 7 Jahren Aufenthalt (finanzielle Zuständigkeit des Bundes) nicht berücksichtigt, eine durchaus substantielle Bevölkerungsgruppe. Hierfür bestehen separate Statistiken (Sozialhilfeempfängerstatistik, Unterstatistiken **SH-FlüStat** und **SH-AsylStat**).

SH-FlüStat steht ab 2010, SH-AsylStat SH erst ab 2016 zur Verfügung. Die Qualität ist in den ersten Jahren bei SH-FlüStat dabei unbefriedigend, wobei ab 2012, welches wir als Startjahr für die Analyse von Zeitreihen vorsehen, jedoch immer über 90% der Dossiers über die SV-Nr. mit STATPOP verknüpfbar sind (Guggisberg 2020 et al., technischer Hintergrundbericht). Zu beachten ist bei der Sozialhilfestatistik, dass diese aktuell grundlegend modernisiert wird und voraussichtlich ab 2025 in einer modifizierten Form zur Verfügung stehen wird.

Wir schlagen vor, die Datensätze SH-FlüStat und SH-AsylStat zur Erzielung einer vollständigeren Abdeckung bezüglich Sozialhilfebezug für den verknüpften Datensatz neu einzubeziehen.

■ Bedarfsabhängige Sozialleistungen im Rahmen der **Alimentenbevorschussung ALBV** werden in WiSiER nicht berücksichtigt (respektive steuerlich gleichbehandelt, weshalb sie – sofern in der Steuererklärung ausgewiesen – nicht unterscheidbar von regulären Unterhaltsbeiträgen sind). Mit dem Wegfall der Steuerdaten fallen diese Informationen automatisch weg. Im Rahmen der Sozialhilfestatistik wird der Bezug dieser Sozialhilfeleistung, welche v.a. für viele einkommensschwache Alleinerziehende eine relevante Einkommenskomponente ist, erfasst. Sie ist zusätzlich auch als bedarfsabhängige Leistung mit Blick auf das Schwerpunktthema materielle Existenzsicherung von Interesse.

■ Das **Familienzulagenregister** der ZAS erfasst seit 2011 Kinder- und Ausbildungszulagen. Erfasst wird dabei, wer von welcher Kasse Zahlungen erhält, nicht aber deren Höhe. Die Betragshöhe wäre entsprechend zu modellieren basierend auf Informationen zu den kantonalen Familienzulagen – analog wie wir dies bereits für die Kinderzulagen bei der Plausibilisierung vornehmen. Der konkrete Vorteil dieser Datenquelle läge denn auch weniger bei den Kinderzulagen, auf die grundsätzlich ein allgemeiner Anspruch besteht und von denen deshalb pragmatisch angenommen werden kann, dass diese bezogen werden,⁴ als vielmehr bei den **Ausbildungszulagen**, auf die nur Kinder in Ausbildung Anspruch haben (ab 15 Jahren bis längstens bis Vollendung des 25. Lebensjahres). Mit dem Familienzulagenregister lassen sich diese Kinder in Ausbildung mit Ausbildungszulagen identifizieren. Der Umgang mit Spezialfällen, bspw. Eltern mit mehrfachem Bezug von Familienzulagen aufgrund Erwerbstätigkeit in mehreren Kantonen (sog. «Differenzzahlung»), wäre bei der konkreten Umsetzung zu klären. Verknüpfungen dieser Daten und entsprechende Analysen wurden bereits innerhalb des BSV und für einzelne externe Forschungsprojekte vorgenommen.

Grundsätzlich scheint ein Einbezug dieser Daten aus Sicht des BASS grundsätzlich auf lange Frist sinnvoll, wobei aufgrund des **relativ geringen Betrags** (insbesondere bezogen auf das Äquivalenzeinkommen einer Person) jedoch fraglich ist, ob der Mehrwert der Informationen bezüglich der Ausbildungszulagen⁵ den Aufwand tatsächlich wert ist. Bei der Abwägung sind die konkret zu verfolgenden Fragestellungen in den Analysen zu berücksichtigen.

■ Angaben zu den **Prämienverbilligungen** sind in den EFIND-Daten des BAG zur Berechnung des Risikoausgleichs ab dem Jahr 2019 ableitbar, da dort **Brutto- wie auch Nettoprämienzahlungen**

⁴ Anspruch haben auch Nichterwerbstätige, sofern sie keine Ergänzungsleistungen beziehen und je nach Kanton eine bestimmte Einkommensobergrenze nicht überschreiten.

⁵ Kinderzulagen sind modellierbar und somit bereits verfügbar.

der obligatorischen Grundversicherung für die gesamte Bevölkerung erfasst werden (telefonische Abklärungen bei Frau Holzgang-Rappo, BFS). Aus der Differenz zwischen Brutto- und Nettoprämie lässt sich die Höhe der ausbezahlten kantonalen Prämienverbilligung ableiten. Für den verknüpften Datensatz ist angesichts der aktuell zahlreichen weiteren fehlenden Einkommenskomponenten beim verfügbaren Einkommen abzuwägen, inwiefern sich für die Informationen zur Prämienverbilligung der Mehraufwand einer Verknüpfung auszahlt. Da der Prämienverbilligung im Rahmen des verfügbaren Einkommens **entsprechende obligatorische Ausgaben in Form der Prämien der obligatorischen Grundversicherung** gegenüberstehen, müssten diese Ausgaben bei Einbezug der Prämienverbilligung ebenfalls konzeptuell in die Einkommensdefinition einbezogen werden, wie dies beim verfügbaren Einkommen auch getan wird. Ein Einbezug ist auch deshalb wichtig, weil die Prämienbelastung von Kanton zu Kanton stark variiert, und ansonsten die Vergleichbarkeit zwischen den Kantonen eingeschränkt ist.

Als Indikator für Einkommensschwäche ist der IPV-Bezug weniger geeignet: Die Prämienverbilligung ist eine Sozialleistung mit zwei Pfeilern mit jeweils stark unterschiedlichem Leistungsumfang und Anspruchskreis. Anspruch haben zum einen Versicherte in bescheidenen wirtschaftlichen Verhältnissen, wobei der Umfang der Leistungen einkommensabhängig ist (Beziehende von Sozialhilfe und Ergänzungsleistungen haben in der Regel automatisch Anspruch). Zum anderen sind die Kantone verpflichtet, bei Familien mit unteren und mittleren Einkommen die Prämien der Kinder und jungen Erwachsenen in Ausbildung um mindestens 80, respektive 50 Prozent zu verbilligen. Der Anspruchskreis bei diesem zweiten Pfeiler reicht weit über die einkommensschwache Bevölkerung hinaus bis in mittlere Einkommensschichten und je nach Haushaltskonstellation und Kanton bis zu jährlichen steuerbaren Haushaltseinkommen von über 100'000 CHF.

Effektives Anspruchsjahr vs. Valuta-Jahr: Erst eine vertiefte Analyse dieser Daten würde zeigen, wie stark die Diskrepanz zwischen Anspruchsjahr der Prämienverbilligung und faktischem Zahlungsjahr eine Herausforderung bei der Verwendung dieser Daten darstellt. Bekannt ist, dass IPV-Anspruchsberechtigte die Prämienverbilligung je nach kantonaler Regelung im laufenden Anspruchsjahr erst teilweise erhalten, oder aber es erfolgen Rückverrechnungen in späteren Jahren (nach Vorliegen der definitiven Steuerveranlagung, mit welcher die Anspruchsberechtigung abschliessend bestimmt wird).

Der Umstand, dass die EFIND-Daten **erst für die Jahre ab 2019 verfügbar** sind, schmälert aktuell noch den potentiellen Mehrwert – sie wären für Verlaufsanalysen deshalb noch nicht einsetzbar. Mit Blick auf eine zukünftige kombinierte Verwendung mit den Steuerdaten und bei Vorhandensein einer längeren Zeitreihe verfügt diese Datenquelle aber längerfristig über grosses Potenzial.

BASS schlägt vor diesem Hintergrund vor, auf die Verknüpfung dieser EFIND-Daten (Prämienverbilligung, KK-Prämien) im Moment noch zu verzichten, mittelfristig aber deren Einbezug vorzusehen.

■ Die **Statistik der kantonalen Stipendien und Darlehen** (STIP, Stipendienregister) des BFS erfasst die entsprechenden Individualdaten mit entsprechender SV-Nr. prinzipiell seit 2010, wobei die Kantone unterschiedlich schnell die Übermittlung der SV-Nr. umgesetzt haben. Seit ca. 2015 verfügen über 95 Prozent der Daten über diese Information. Stipendiendaten wurden bereits im Rahmen von externen Forschungsprojekten mit weiteren Administrativdaten verknüpft und entsprechend ausgewertet (laufendes Projekt der Berner Fachhochschule, Hümbelin et al. 2021).

Von den Studierenden werden 7.4 Prozent mit im Schnitt rund 7'500 CHF unterstützt (BFS – STIP 2021), allerdings bilden die Studierenden selber eine relativ kleine Gruppe innerhalb der Bevölkerung. Deren finanzielle Verhältnisse sind generell schwierig korrekt abbildbar, da private Transferzahlungen durch die Eltern bei einem beträchtlichen Teil der Studierendengruppe, der ausserhalb

des Elternhaushaltes wohnt und nicht stipendienberechtigt ist, eine wichtige Einkommensquelle darstellen. Im Rahmen von WiSiER werden deshalb Haushalte (nur) von Personen unter 25 Jahren separat und eingeschränkt analysiert (vgl. etwa Wanner/Gerber 2021: 55f). Beim Einbezug von Stipendienleistungen bei einem Indikator wäre vor diesem Hintergrund unbedingt zu prüfen, wie mit diesen Haushalten umgegangen wird, damit nicht fälschlicherweise stipendienbeziehende Studierende als finanziell bessergestellt als Studierende, die von den Eltern mit privaten (und nicht in den Daten erfassten) Transferzahlungen unterstützt werden, erscheinen.

Vor diesem Hintergrund (kleiner Anteil Betroffener, generelle Schwierigkeit der Bestimmung der finanziellen Situation von Studierenden) scheint ein Einbezug dieser Daten von nachgeordneter Priorität.

■ Die Daten der **Schweizerischen Gesundheitsbefragung SGB** des BFS sind zentral für gesundheitliche Fragestellungen. Die Schweizer Gesundheitsbefragung erfasst detailliert Angaben zur **Gesundheit**, aber auch etablierte Items zu sozialer Unterstützung (Aspekt **soziale Beziehungen** des Armutsmonitorings). Sie wird alle 5 Jahre durchgeführt und befragt jeweils rund 22'000 Bewohner und Bewohnerinnen, jüngst im Jahr 2022. Angaben zum sozioökonomischen Status, inklusive dem Haushaltseinkommen, werden in der SGB selber bereits abgefragt. Die SGB bietet deshalb grosses Potenzial zur Beantwortung entsprechender Fragestellungen. Das BASS hat in einer früheren Studie (Guggisberg, Höglinger et al. 2022) SGB-Daten mit STATPOP und IK-Daten verknüpft, um die Erwerbs- und Einkommenssituation der Befragten akkurater abzubilden).

Aus Sicht des BASS ist ein Einbezug der Schweizerischen Gesundheitsbefragung von nachgeordneter Priorität. Dies deshalb, weil die SGB-Daten mit ihrer vergleichsweise kleinen Querschnittsstichprobe und dem mehrjährigen Erhebungszyklus nur sehr limitiert in übergreifende Analysevorhaben (inkl. Verlaufs- und kleinräumige Analysen) eingebaut werden könnten. Geht es hingegen einzig darum, den Zusammenhang zwischen Gesundheit und finanzieller Situation auszuleuchten, so stehen bereits im SGB (selbst-reportierte) Angaben zur Verfügung oder wäre eine Verknüpfung (nur) mit den IK-Daten eine zielführende und weitaus weniger aufwändige Möglichkeit.

■ Die **Schweizerische Lohnstrukturerhebung LSE** ist eine bei den Betrieben durchgeführte Erhebung mittels Stichprobenverfahren, die alle 2 Jahre durchgeführt wird. In der jüngsten publizierten Runde von 2020 wurden Daten von rund 2,1 Mio. Erwerbstätigen erfasst (bei gesamthaft etwas über 5 Mio. Erwerbstätigen). Die LSE erfasst u.a. auch Angaben zu Branche, Beruf und insbesondere **Beschäftigungspensum**. Letzteres wäre eine äusserst wertvolle Information für zahlreiche Analysevorhaben zum Erwerbseinkommen, bei welchen die Unterscheidung bei tiefem Erwerbseinkommen als Folge von Tieflohn vs. tiefem Arbeitspensum höchst relevant ist, was allein anhand der IK-Daten nicht möglich ist.

Bei einem Einbezug der LSE-Daten ist zu berücksichtigen, dass diese zumindest in früheren Jahren nicht oder nur begrenzt plausibilisiert wurden und die Daten je nach spezifischer Fragestellung deshalb nur bedingt geeignet sind. Das BASS hat im Rahmen von früheren Projekten auch entsprechende Plausibilitäts-Filter für die LSE-Daten erstellt, die man entsprechend verwenden könnte. Da die Erhebungseinheit Unternehmen sind und die vertragliche Arbeitszeit erfasst wird, stellen insbesondere Personen mit mehreren Anstellungen sowie atypisch Beschäftigte (u.a. Stundenlohn, unregelmässige Pensen) eine Herausforderung bei der Analyse dar. Angestellte von Unternehmen mit weniger als 3 Beschäftigten sowie Selbständige werden zudem nicht erfasst. Für Verlaufsanalysen einschränkend ist auch der nur 2-jährige Erhebungsrhythmus sowie die (allerdings sehr grosse)

Stichprobe. Die SV-Nr. ist im relevanten Datenzeitraum ab 2012 verfügbar und eine Verknüpfung somit grundsätzlich möglich.

Als **alternative Quelle bezüglich Arbeitszeit** bietet sich eine wenig verwendete Variable in der **Strukturerhebung SE** des BFS an. Dort wird die Anzahl wöchentlicher Arbeitsstunden erfragt, auf Basis deren das Arbeitspensum berechnet werden kann. Vorteilhaft ist, dass dieses nicht nur für die Zielperson, sondern für alle Erwerbstätigen eines Haushalts erfragt wird. In der Strukturerhebung ist die Angabe im Gegensatz zur LSE nicht die vertragliche Arbeitszeit, sondern die selbstdeklarierte tatsächlich geleistete Arbeitszeit. Ein Vergleich dieser Variable mit den entsprechenden LSE-basierten Informationen wurde bereits vom BFS im Rahmen von LABB-Analysen durchgeführt. Konkret gibt es gewisse Abweichungen bei den Ergebnissen, wobei sich diese in engen Grenzen halten. Dabei ist keine der beiden Datenquellen besser, die Unterschiede dürften vorwiegend das Ergebnis der konzeptionellen Differenzen sein (vgl. BFS 2021). Ein grösserer Vorteil für unsere Zwecke ist, dass diese Information bei der SE wie erwähnt jeweils für **alle Erwerbstätigen des gesamten Haushalts** zur Verfügung steht, da dies auch für die weiteren Haushaltsmitglieder erfragt wird. Somit ist für Haushalte in der Stichprobe diese Information vollständig verfügbar und es kann das Ausmass der Erwerbstätigkeit auf Haushalts-Ebene bestimmt werden. Mit den Angaben aus der LSE ist dies nur eingeschränkt möglich.

Bei beiden Datenquellen, SE und LSE, ist zu beachten, dass sie keine Vollerhebungen sind und deshalb **Längsschnittanalysen mit diesen zeitvarianten Variablen deshalb nur begrenzt möglich** sind. Dies schmälert den Nutzen dieser Variable Arbeitszeit leider wieder ein.

Nach Prüfung der beiden Varianten schlagen wir vor, diese Information zur Arbeitszeit primär aus der bereits verfügbaren Strukturerhebung zu verwenden und auf den zusätzlichen Einbezug und die Aufbereitung dieser Information aus den LSE-Daten vorerst zu verzichten (oder nur als sekundäre Option vorzusehen).

3 Abdeckung der relevanten Einkommenskomponenten mit dem verknüpften Datensatz

In den folgenden beiden Unterkapiteln wird ausgeführt, welche relevanten Einkommenskomponenten zur Bestimmung des Haushaltseinkommens im verknüpften Datensatz (ohne Steuerdaten) verfügbar sind und welche nicht (Kapitel 3.1). Und es wird auf spezifische Bevölkerungsgruppen eingegangen, deren Einkommen auch mit dem WiSiER-Datensatz bereits nur unvollständig abgedeckt werden konnten, konkret AHV-Selbständige, Quellenbesteuerte (Kapitel 3.2).

3.1 Verfügbare und fehlende Einkommenskomponenten im verknüpften Datensatz

Die Angaben im vorgesehenen verknüpften Datensatz ermöglichen – wie auch bereits bei WiSiER – eine Zusammenführung der Einkommen unterschiedlicher Personen eines Haushalts, um das für die meisten Analysen relevante **Haushaltseinkommen** zu bestimmen (Haushaltsbildung).

■ Das **Erwerbseinkommen** stellt für die überwiegende Mehrheit der Haushalte im erwerbsfähigen Alter die Haupteinnahmequelle dar. Angaben dazu sind im verknüpften Datensatz auf Basis der IK-Daten vorhanden und überwiegend von guter Qualität.⁶

Gewisse Einschränkungen bestehen bei den AHV-Selbständigen, wobei die Bemessung des Erwerbseinkommens dieser Gruppe generell eine Herausforderung darstellt, unabhängig von der dafür verwendeten Datenquelle.

■ Weiter sind Angaben zu den **Sozialversicherungsleistungen AHV und IV** vorhanden.

■ Auch sind die **wichtigsten bedarfsabhängigen Sozialleistungen** enthalten, namentlich die **wirtschaftliche Sozialhilfe** und die **Ergänzungsleistungen**.

Unterschiede gegenüber dem WiSiER-Datensatz

Aufgrund der fehlenden kantonalen Steuerdaten fehlen gegenüber den verknüpften WiSiER-Daten einige wesentliche Einkommenskomponenten, die im Folgenden ausgeführt werden. Gleichzeitig besteht mit der Integration von neuen Datenquellen die Möglichkeit, gewisse in WiSiER bestehende Lücken zu schliessen. Der Fokus liegt dabei auf jenen Einkommenskomponenten, die generell oder für bestimmte relevante Bevölkerungsgruppen/Haushaltstypen einen wesentlichen Einkommensbestandteil bilden.

Gegenüber WiSiER fehlende relevante Einkommenskomponenten:

■ **Unterhaltszahlungen (erhaltene, sowie auch ausgerichtete)**. Unterhaltszahlungen für minderjährige Kinder, nicht aber volljährige Kinder in Ausbildung, sind in den Steuerdaten erfasst und somit in WiSiER vorhanden, sofern korrekt angegeben.⁷ Alimentenzahlungen sind ein potentiell wesentlicher Einkommensbestandteil für Einelternerhaushalte und, im Gegenzug, schmälern sie das verfügbare Einkommen von Haushalten von unterstützungspflichtigen Elternteilen. Ohne kantonale Steuerdaten fehlen Angaben dazu bis auf weiteres.

■ **Renten der beruflichen Vorsorge BVG**. Für Haushalte ohne Personen im Rentenalter, die relevante Grundgesamtheit, sind Renteneinkünfte nur für eine Minderheit relevant, die es aber im Auge zu behalten gilt. Ein potentiell wesentlicher Einkommensbestandteil sind Renten der beruflichen

⁶ Die Umrechnung des in den IK-Daten erfassten Bruttoeinkommens in das Nettoeinkommen kann unter Berücksichtigung der Sozialversicherungsabzüge einfach modelliert werden.

⁷ Unterhaltszahlungen an volljährige Kinder in Ausbildung sind steuerlich nicht relevant und deshalb auch nicht in den Steuerdaten erfasst. Hingegen sind auch Alimente an geschiedene oder getrennte Ehepartner oder Ehepartnerinnen steuerlich relevant.

Vorsorge für Haushalte von IV-Rentenbeziehenden sowie, von weitaus geringerer Zahl, Haushalte von Hinterlassenen (Witwen-, Witwer- und Waisenrenten). Renten der beruflichen Vorsorge sind ebenfalls ein wesentlicher Einkommensbestandteil bei älteren Personen vor Erreichen des ordentlichen Rentenalters, welche einen Rentenvorbezug aus der 2. Säule tätigen. (Vorhanden sind hingegen auch im verknüpften Datensatz die Angaben zu Renten der 1. Säule AHV/IV, inkl. Hinterlassenleistungen, sowie zu den Ergänzungsleistungen.)

Gegenüber WiSiER fehlende Einkommenskomponenten von sekundärer Relevanz sind:

■ **Vermögen und Vermögensverzehr.** Das Einkommen aus Vermögen ist von geringer Relevanz für das Armutsmonitoring. Stärker ins Gewicht fällt die ohne Angaben zum Vermögen fehlende Möglichkeit der Berechnung eines allfälligen Vermögensverzehrs, wie dies etwa im Rahmen von WiSiER gemacht wurde. Darauf hinzuweisen ist, dass bei den etablierten Armutsindikatoren des BFS basierend auf den SILC-Daten das Vermögen ebenfalls nicht berücksichtigt wird, obwohl dies grundsätzlich sinnvoll wäre.⁸

■ **Krankentaggelder.** In den IK-Daten sind diese nicht erfasst, d.h. während dieser Phase ist kein Lohn oder ein stark reduzierter AHV-pflichtiger Lohn ausgewiesen (letzteres dann, wenn Arbeitgebende die 20% Differenz des Krankentaggelds zum Lohn bezahlen). In den kantonalen Steuerdaten wären diese Informationen vorhanden.

■ **Kinder- und Ausbildungszulagen.** Kinderzulagen bis zum 16. Lebensjahr sind relativ einfach befriedigend modellierbar, dies gilt jedoch nicht für Ausbildungszulagen für erwachsene Kinder bis 25 Jahre, da Angaben zum Ausbildungsstatus der volljährigen Kinder fehlen.

Gegenüber WiSiER (teils) verfügbare relevante Einkommenskomponenten:

Auch im WiSiER-Datensatz fehlen relevante Einkommenskomponenten. Mit dem Einbezug zusätzlicher individueller Datenquellen könnten diese Lücken bei Bedarf teilweise geschlossen werden. Mit Stern (*) markiert sind jene Komponenten, die mit den oben in Kapitel 2.2 aufgeführten Datenquellen grundsätzlich abgedeckt werden könnten.

■ **weitere bedarfsabhängige Sozialleistungen** (neben den Ergänzungsleistungen und der wirtschaftlichen Sozialhilfe, für welche die Informationen bereits vorhanden sind):

- **Sozialhilfe im Asyl- und Flüchtlingsbereich***

(Aufenthalt unter 5 Jahre bei anerkannten Flüchtlingen, respektive 7 Jahre bei vorläufig aufgenommenen Flüchtlingen und vorläufig aufgenommenen Personen VAP)

- **Alimentenbevorschussung ALBV***

- **Ausbildungsstipendien***

- **Prämienverbilligung IPV***

- allfällige kantonale Familienbeihilfen (u.a. «Familien-EL») und Beihilfen zur EL, kommunale Bedarfsleistungen.

■ **obligatorische Ausgaben**, wie sie für die Bestimmung des verfügbaren Einkommens relevant sind.

- **Prämien der obligatorischen Krankenversicherung***

- **Wohnkosten** (Miete oder Kosten des selbstbewohnten Wohneigentums). Dieser Ausgabenposten bestimmt das verfügbare Einkommen stark und es wäre mit Blick auf einen zukünftigen Datensatz

⁸ Entsprechende ergänzende Konzepte/Indikatoren zur Messung von Armut mittels SILC-Daten unter Einbezug des Vermögens sind jedoch beim BFS in Entwicklung und werden geprüft (vgl. [Link](#)).

mit kantonalen Steuerdaten zu prüfen, inwiefern auch diese relevante Lücke zur Bestimmung des verfügbaren Einkommens geschlossen werden kann.

Eine tabellarische Gegenüberstellung von WiSiER (mit Steuerdaten) einerseits und dem vorgesehenen verknüpften Datensatz andererseits bezüglich vorhandener und nicht vorhandener Informationen zu allen Einkommenskomponenten zur Bestimmung des verfügbaren Einkommens (basierend auf dem idealtypischen Einkommenskonzept des Canberra Group Handbooks, UN ECE 2011) ist als **Beilage zur vorliegenden Studie in Form einer Excel-Tabelle** vorhanden.

3.2 Einkommensinformationen zu spezifischen Personengruppen

Mit dem Verzicht auf die Nutzung kantonalen Steuerdaten wird der Ausschluss von quellenbesteuerten ausländischen Personen hinfällig. Weiterhin bestehen bleibt die Schwierigkeit der adäquaten Abbildung der AHV-Selbständigen.

Erwerbseinkommen von Quellenbesteuerten (Vollständigkeit bezüglich ausl. Bevölkerung)

Der Umstand, dass ausländische Personen ohne Niederlassungsbewilligung unterhalb einer definierten Einkommensobergrenze (rund 120'000 CHF) i.d. R. quellenbesteuert werden und deshalb für diese keine regulären Steuerinformationen verfügbar sind, hat zur Folge, dass diese Gruppe in WiSiER-basierten Analysen meist nur begrenzt berücksichtigt, respektive ausgeschlossen wurden. Konkret handelt es sich um Ausländerinnen und Ausländer mit Aufenthaltsbewilligung B, aber auch Personen aus dem Asyl- und Flüchtlingsbereich (Aufenthaltsstati F, N, Schutzstatus S). Durch die Abstützung auf die IK-Daten anstelle der Steuerdaten entfällt für den verknüpften Datensatz diese Schwierigkeit. **Die ausländische Bevölkerung ist somit umfassend abgebildet, ein wesentlicher Vorteil des verknüpften Datensatzes**, der sich nicht auf die Steuerinformationen abstützt.

Erwerbseinkommen von AHV-Selbständigerwerbenden

Die finanzielle Situation von Selbständigen ist grundsätzlich schwierig zu bestimmen, da diese über einen beträchtlichen Spielraum bei der Ausweisung ihres Einkommens verfügen (vgl. Wanner 2019: 18f.). Basierend auf dem von den Steuerbehörden berechneten Einkommen wird AHV-Selbständigen für die IK ein bereinigtes Einkommen berechnet, welches aufgrund von Abzugsmöglichkeiten massig tiefer als das tatsächliche erzielte Einkommen liegt. Auch geschieht der Eintrag in den IK-Daten mit einer längeren Verzögerung von bis zu 3-4 Jahren gegenüber Arbeitnehmenden. Für Selbständige, die nur den AHV-Mindestbeitrag entrichten, wird ein fiktives Einkommen zwischen 9'000 und 9'500 CHF in den IK gutgeschrieben. Zu berücksichtigen ist, dass die AHV-Selbständigkeit eng definiert wird. Diese umfasst nur Arbeit auf eigene Rechnung, üblicherweise in einem Einzelunternehmen, nicht aber Selbständigkeit in Form von «Arbeitnehmenden» im eigenen Betrieb (GmbH, AG). Zu berücksichtigen ist weiter, dass eine grössere Zahl an Selbständigen zusätzlich noch ein Erwerbseinkommen aus einer Anstellung beziehen (vgl. BSV 2023b).

4 Plausibilisierung der verknüpften Daten ohne Steuerinformationen zur Bestimmung der einkommensschwachen Bevölkerung

Mit der im folgenden vorgenommenen Plausibilisierung soll aufgezeigt werden, mit welcher Genauigkeit die einkommensschwache Bevölkerung mit den zukünftig verfügbaren verknüpften Administrativ- und Befragungsdaten (ohne Steuerdaten) identifiziert werden kann. Im folgenden Abschnitt 4.1 wird das Vorgehen detailliert ausgeführt, die Ergebnisse werden in Abschnitt 4.1.2 präsentiert und diskutiert.

4.1 Vorgehen bei der Plausibilisierung

Das Vorgehen ist dabei in aller Kürze wie folgt: Als Indikator für die einkommensschwache Bevölkerung dient der Indikator P20 – die 20 Prozent der Bevölkerung mit dem tiefsten Haushaltsäquivalenzeinkommen. Sämtliche Analysen werden mit den WiSiER-Daten durchgeführt, wobei der Indikator berechnet basierend auf WiSiER inkl. Steuerdaten (im Folgenden und in den Auswertungen «Referenzdatenbasis WiSiER mit Steuer» genannt)⁹ den Ergebnissen auf Basis der verknüpften Daten ohne Steuerdaten gegenübergestellt werden. Auf Basis der verknüpften Daten ohne Steuerdaten wird der Indikator dabei in drei leicht unterschiedlichen Varianten berechnet: Der Ausgangsvariante «verknüpfte Daten ohne Steuer», einer optimierten Variante sowie einer Variante unter ausschliesslicher Verwendung des Erwerbseinkommens (sowie allfälliger ALV-Leistungen).

4.1.1 «Einkommensschwäche» P20 als primär verwendeter Indikator

Die Plausibilisierung wird primär für den Indikator der **Einkommensschwäche P20** – die untersten 20 Prozent der Bevölkerung mit dem tiefsten Haushaltsäquivalenzeinkommen – durchgeführt (siehe Textkasten). Zusätzlich erfolgt eine Plausibilisierung auch mit dem alternativen Grenzwert von 60 Prozent des medianen Haushaltseinkommens («**60% Median**»), wie er auch für das breit etablierte relative Armutsmass der Armutsgefährdungsgrenze zur Anwendung kommt.

4.1.1.1.1.1 Die «einkommensschwache Bevölkerung» bestimmt als P20

Mit dem verknüpften Datensatz sollen ergänzende Analysevorhaben im Rahmen des Armutsmonitoring durchgeführt werden. Dabei ist zu berücksichtigen, dass das verfügbare Einkommen mit den verknüpften Daten – primär aufgrund fehlender Steuerdaten, aber auch wegen weiterer fehlender Komponenten – nur lückenhaft bestimmt werden kann. Die geplante Vorgehensweise vor diesem Hintergrund ist es, zur Abgrenzung gegenüber den im Rahmen des Armutsmonitorings andernorts verwendeten Armutskonzepten ein neues, alternatives Konzept einzuführen: Die **«einkommensschwache Bevölkerung»** (vgl. Bio Tchané 2019, Fluder et al. 2020), bestimmt als die untersten 20 Prozent der Bevölkerung mit dem tiefsten Haushaltsäquivalenzeinkommen.

Ein Vorteil dieser Kennzahl ist seine **Einfachheit** und **(relative) Voraussetzungslosigkeit**, was im Kontext von Daten mit Limitationen attraktiv ist. So benötigt das P20-Konzept lediglich eine annähernd korrekte Reihung der Haushalte im Bereich des 20er-Perzentils, während etwa das ebenfalls relative Konzept der «Armutsgefährdungsgrenze» (60% des medianen Haushaltseinkommens) zusätzlich auch eine annähernd korrekte substantielle Bestimmung des Haushaltseinkommens, und zwar bis zum 50er-Perzentil, das als Referenz dient, voraussetzt. Im Hinblick auf das

⁹ Dabei ist auch eine kritische Beurteilung der verknüpften Steuerdaten vorzunehmen, die zwar eine vergleichsweise qualitativ hochstehende, aber ebenfalls nicht fehlerlose Bestimmung des Einkommens ermöglichen (bspw. besteht die Möglichkeit von in der Steuerdeklaration nicht angegebenen Einkommen, von in der kalenderjährlichen Steuerperspektive nicht abgebildeten unterjährigen Dynamiken, von auch hier fehlenden Einkommenskomponenten usw.).

Armutsmonitoring ist zudem wichtig, dass eine klare (auch begriffliche) Abgrenzung zu anderen im Armutsmonitoring verwendeten Armutskonzepten vorgenommen wird (wie etwa absolute Armut/Armutsquote, Armutsgefährdung).

Das BFS verwendet in Analysen zur «Einkommensmitte» den Begriff «Einkommensschwach» zur Abgrenzung der Bevölkerung unterhalb der «Mitte» mit weniger als 70% des medianen Haushaltseinkommens ([Link](#)). Vom Umfang deckt sich die so definierte einkommensschwache Bevölkerung empirisch grob mit der Gruppe der P20. Generell ist «Einkommensschwäche» ein in der Literatur verwendeter Begriff für Personen innerhalb der Einkommensverteilung unterhalb einer bestimmten Grenze (vgl. etwa Drilling 2004), ohne dass es eine fest etablierte Definition gibt.

Aufgrund der rein relativen Bestimmung der P20 ist per Definition der Anteil der einkommensschwachen Bevölkerung an der Grundgesamtheit unveränderlich. Somit ist das Konzept nicht geeignet, um die Entwicklung der Armut auf Ebene der Gesamtbevölkerung zu beobachten (Zunahme/Abnahme über Zeit), wie dies etwa die Konzepte absolute Armut oder Armutsgefährdung erlauben. Ein weiterer Nachteil der rein relativen Bestimmung ist, dass fehlerhafte Bestimmungen von Personen oder Personengruppen immer auch Auswirkungen auf die Reihung insgesamt haben. Das gleiche gilt für allfällige vorgenommene Ausschlüsse (etwa aus Qualitätsgründen), die ebenfalls auf die gesamte Reihung und somit auf die Ergebnisse Auswirkungen haben.

Vergleichbares Haushaltseinkommen mittels neuer OECD-Skala

Die Indikatoren werden dabei auf Haushaltsebene unter Verwendung des Äquivalenz-Haushaltseinkommens berechnet, die Auswertung bezüglich des Anteils der Betroffenen erfolgt dann aber auf individueller Ebene der Personen. Damit folgen wir der gängigen Praxis wie Armutsindikatoren ausgewiesen werden, etwa vom BFS.

Die Einkommenssituation wird gemäss gängiger Praxis auf der Haushaltsebene bestimmt, da Haushalte als Einkommens- und Konsumgemeinschaft fungieren. Voraussetzung für die Bestimmung des Haushaltseinkommens ist die Möglichkeit der Haushaltsbildung, d.h. die Zusammenführung der individuellen Angaben der Haushalts-Mitglieder. Mit den verfügbaren Daten (STATPOP) ist eine **Haushaltsbildung für den Zeitraum ab 2012 möglich**.¹⁰ Für die Analysen wird die für die WiSiER-Daten erarbeitete Haushaltstypologie verwendet (vgl. Wanner 2019: 48-51, Wanner/Gerber 2021: 15f).

Um den Vergleich zwischen verschiedenen Typen von Haushalten zu ermöglichen, wird die modifizierte oder **neue OECD-Äquivalenzskala für das Haushaltseinkommen** verwendet. Dabei geht der oder die Hauptbeziehende des Einkommens mit dem Faktor 1.0 in die Gewichtung ein, alle weiteren Mitglieder des Haushaltes ab 14 Jahren erhalten den Faktor 0.5, jüngere den Faktor 0.3. Bei der modifizierten OECD-Skala handelt es sich um eine etablierte Expertenskala, die primär für statistische Vergleichszwecke von Haushalten über alle Einkommensschichten hinweg entwickelt wurde. Sie wird auch bei den anderen Armutsmassen im Rahmen des Armutsmonitorings verwendet.¹¹

¹⁰ Bei den STATPOP-Daten der Jahre 2010-2011 fehlen die entsprechenden Haushalts-Identifikatoren bei den Personen.

¹¹ Allfällige Analysen, welche den Zeitraum vor 2012 einbeziehen, müssen mit der Tatsache umgehen, dass hierfür keine Angaben zum Haushalt, sondern lediglich individuelle Daten zur Verfügung stehen (v.a. Erwerbseinkommen und -verläufe basierend auf IK-Daten), eine Haushaltsbildung ist nicht möglich.

Mit den verfügbaren Daten ist weiter auch mehrheitlich, aber nicht für alle Bevölkerungsgruppen in gleich gutem Ausmass die Zuordnung von Personen zu ihren Eltern möglich, und damit auch zur Erwerbsbiographie der Eltern (Häufiger Schwierigkeiten diesbezüglich sind bei Personen mit Migrationshintergrund zu erwarten). Die entsprechenden Ergebnisse für individuelle Personen können dann wiederum mit den Haushalten, in denen diese Personen ab 2012 leben, verbunden werden. Bspw. kann dadurch aufgezeigt werden, inwiefern (haushalts-)einkommensschwache Individuen im Zeitraum ab 2012 sich bezüglich ihrer

4.1.2 Die verwendeten Varianten der Berechnung des Haushaltseinkommens

Für die Plausibilisierung wird das Haushaltseinkommen immer unter Verwendung der WiSiER-Datenbasis berechnet, allerdings jeweils unterschiedlich: Als Referenzgrösse unter Verwendung von kantonalen Steuerinformationen. Für das zu plausibilisierende Haushaltseinkommen ohne Steuerdaten werden drei Varianten berechnet: Die Standardvariante ohne Steuerdaten, eine optimierte Variante davon sowie eine auf das Erwerbseinkommen fokussierte Variante.

Haushaltseinkommen «WiSiER mit Steuer» (Referenzdatenbasis)

Das WiSiER-Haushaltseinkommen berechnet mithilfe der Steuerdaten dient als Referenz, mit dem die weiteren Berechnungsarten ohne Steuerdaten jeweils verglichen werden. Für die Plausibilisierung wird dabei das gängige WiSiER-Haushaltseinkommen berechnet, jedoch ohne den bei WiSiER-Analysen üblicherweise zusätzlich einberechneten Verzehr von 5 Prozent des rasch verfügbaren Vermögens (vgl. Wanner/Gerber 2021: 4). Dies deshalb, da auch bei den üblichen Armutsindikatoren basierend auf SILC-Daten das Vermögen nicht mitberücksichtigt wird. Damit wird eine zusätzliche Quelle von Unterschieden zwischen den beiden Datenquellen ausgeschlossen.

Haushaltseinkommen «verknüpfte Daten ohne Steuer» (Variante 1)

Das Haushaltseinkommen «verknüpfte Daten ohne Steuer» basiert auf den verknüpften Administrativ- und Befragungsdaten und den dort verfügbaren Einkommenskomponenten, wie sie auch zukünftig verfügbar sein werden. Da anders als beim WiSiER-Haushaltseinkommen keine Steuerinformationen mehr vorliegen werden, müssen die entsprechenden Angaben anderweitig herangezogen werden, respektive fehlen diese. Namentlich das Erwerbseinkommen wird alternativ mittels der Informationen aus den Individuellen Konten (IK-Daten) bestimmt.

Variante Haushaltseinkommen «verknüpfte Daten ohne Steuer (optimiert)» (Variante 2)

Zusätzlich zur Plausibilisierung des Haushaltseinkommens «verknüpfte Daten ohne Steuer» wird auch eine Version des Haushaltseinkommens basierend auf den verknüpften Daten ohne Steuer plausibilisiert, die optimiert wurde und bei welcher die Vergleichbarkeit und Qualität höher sein sollte. Konkret wurden folgende Optimierungen vorgenommen:

■ Beim Erwerbseinkommen wurde das Bruttoerwerbseinkommen basierend auf den IK-Daten in ein **Nettoerwerbseinkommen umgerechnet**, wie es entsprechend bei WiSiER basierend auf den Steuerdaten vorliegt. Dabei wurden die obligatorischen Abzüge in Form von Beiträgen an AHV/IV/EO/ALV berücksichtigt.¹²

Bewusst nicht berücksichtigt sind die Lohnabzüge, respektive die Beitragssätze für die berufliche Vorsorge, da diese je nach Pensionskasse und Finanzierungsart unterschiedlich ausfallen und deshalb nur mit beträchtlicher Unsicherheit modelliert werden können. Dies gilt insbesondere auch für die für die Indikatoren relevanten unteren Einkommensbereiche, namentlich Einkommen unterhalb der gesetzlichen BVG-Eintrittsschwelle.

■ **Modellierung Kinderzulagen:** Das Haushaltseinkommen bei Haushalten mit Kindern bis 16 Jahren wird automatisch um die Höhe der Kinderzulage im entsprechende Wohnkanton ergänzt. Bei kantonal unterschiedlichen Zulagen nach Anzahl oder Alter der Kinder wird jeweils die tiefere Zulage berücksichtigt. Die mittlere Höhe (Median) der kantonalen Kinderzulagen per 1.1.2015 beträgt 200

vorgängigen Erwerbsbiographie und allenfalls auch der Erwerbsbiographie ihrer Eltern vom Rest der Bevölkerung unterscheiden.

¹² Dabei wurden für alle Einkommen die Sätze für Arbeitnehmende angewandt: 5.15% (AHV 4.2%, IV 0.7%, EO 0.25%) sowie Arbeitslosenversicherung: 1.1% für Einkommensteile bis 148'200 CHF/Jahr; 0.5 für Einkommensteile über 148'200 CHF/Jahr.

CHF pro Monat.

Ausbildungszulagen für erwachsene Kinder werden nicht modelliert, da aus den Daten nicht ersichtlich ist, welche Kinder sich in einer nachobligatorischen Ausbildung befinden. Ebenfalls nicht berücksichtigt werden kann, ob aufgrund einer ausserkantonalen Erwerbstätigkeit die Höhe der Kinderzulagen anders ausfällt (inkl. Mehrfach-, respektive Differenz-Beziehende, wenn beide Elternteile erwerbstätig sind). Auch über den möglichen Nicht-Bezug von Kinderzulagen gibt es keine Angaben.

Variante Haushaltseinkommen «nur Erwerbseinkommen» (Variante 3)

Als weitere Variante verwenden wir bei der Plausibilisierung auch allein das Erwerbseinkommen eines Haushaltes. Der Vorteil einer solchen fokussierten Betrachtungsweise ist, dass zum einen das Erwerbseinkommen für die überwiegende Mehrheit der Haushalte von Personen im Erwerbsalter die **primäre Einkommensquelle** darstellt und zum anderen dieses **mittels der verknüpften Daten präzise und lückenlos bestimmt** werden kann. Somit sprechen sowohl konzeptuelle Gründe als auch Validitäts-Argumente für den Einbezug dieser Variante.

Neben dem **Nettolohn** werden bei der Variante Erwerbseinkommen auch allfällige Leistungen der **Arbeitslosenversicherung ALV** mitberücksichtigt, da diese mit dem Erwerb sachlich verknüpft sind und zum Einkommen basierend auf Erwerbstätigkeit im weiteren Sinn zu zählen sind.

Wir sehen diesen möglichen Indikator der Erwerbseinkommensschwäche dabei nicht als Ersatz für einen Armutsindikator, sondern vielmehr als erkenntnisreiche Ergänzung und mögliche pragmatische Übergangslösung, solange eine hinreichend präzise Bestimmung des Haushaltseinkommens aufgrund fehlender Steuerdaten nicht möglich ist.

Diese Variante ist zudem auch konzeptuell **anschlussfähig an Analysen über längere Zeitreihen** (ab frühestens 1988) basierend auf den IK-Daten, für welche ebenfalls nur die Erwerbseinkommen verfügbar sind (hier jedoch nicht auf Haushaltsebene, sondern auf individueller Ebene).

4.1.3 Grundgesamtheit und Auswahl der relevanten Beobachtungen

Wir verwenden für die Plausibilisierung die jüngsten verfügbaren WiSiER-Daten des Jahres 2015. Die interessierende Grundgesamtheit für die Plausibilisierung bildet die **ständige Wohnbevölkerung der Schweiz, die in Privathaushalten lebt**. Nicht berücksichtigt werden Personen in Kollektivhaushalten – stationäre Einrichtungen wie Heime, Gefängnisse) und Personen in Sammelhaushalten – fiktive Haushalte aus statistischen Gründen, denen anderweitig nicht zuordenbare Personen zugeteilt werden.

Einschränkungen der WiSiER Daten

Der WiSiER-Datensatz deckt die interessierende Grundgesamtheit aus verschiedenen Gründen nicht vollständig ab, weshalb sich die Plausibilisierung auf eine klar definierte Untermenge daraus beschränkt.

Der WiSiER-Datensatz umfasst nicht alle, sondern **lediglich elf Kantone** (Bern, Luzern, Nidwalden, Basel-Stadt, Basel-Land, Sankt Gallen, Aargau ohne Aarau, Tessin, Wallis, Neuenburg, Genf), die aber doch als repräsentatives Abbild der gesamten Schweiz betrachtet werden können (Wanner 2019: 5-7).

Quellenbesteuerte (sowie pauschalbesteuerte) Personen werden in den WiSiER-Analysen in der Regel nicht berücksichtigt, da für diese keine regulären Steuerdaten vorliegen. Die Quellenbesteuerung betrifft insbesondere **Personen mit ausländischer Staatsangehörigkeit ohne Niederlassungsbewilligung C** unterhalb einer bestimmten Einkommensobergrenze. Personen mit ausländischer Staatsangehörigkeit ohne Niederlassungsbewilligung, faktisch überwiegend Personen mit

Aufenthaltsbewilligung B aber auch Personen mit Asyl- und Flüchtlingshintergrund, umfassen Ende 2015 knapp 10 Prozent der Schweizer Bevölkerung oder etwa ein Drittel der ausländischen Bevölkerung (Wanner/Gerber 2021: 12). Für die Plausibilisierung werden Haushalte mit Personen ohne Schweizer Staatsangehörigkeit oder ausländische Personen ohne Niederlassungsbewilligung C ausgeschlossen.

Diese Einschränkung der WiSiER-Daten bezüglich der unvollständigen Abdeckung der ausländischen Bevölkerung würde mit der Nutzung von verknüpften Administrativ- und Befragungsdaten (ohne Steuerinformationen) wegfallen, da die entsprechenden IK-Daten für die gesamte Bevölkerung verfügbar sind. Dies ist wie oben bereits erwähnt ein Vorteil der Datengrundlage ohne Steuerinformationen.

Im Rahmen von WiSiER werden zudem **Haushalte von (nur) Personen unter 25 Jahren** separat und eingeschränkt analysiert (vgl. etwa Wanner/Gerber 2021: 55f), da sich deren finanzielle Situation nur unvollständig abbilden lässt. Insbesondere fehlen private Transferzahlungen durch die Eltern, welche bei einem beträchtlichen Teil der Studierendengruppe, der ausserhalb des Elternhaushaltes wohnt und nicht stipendienberechtigt ist, eine wichtige Einkommensquelle darstellen. Diese Haushalte schliessen wir deshalb für die Plausibilisierung ebenfalls aus.

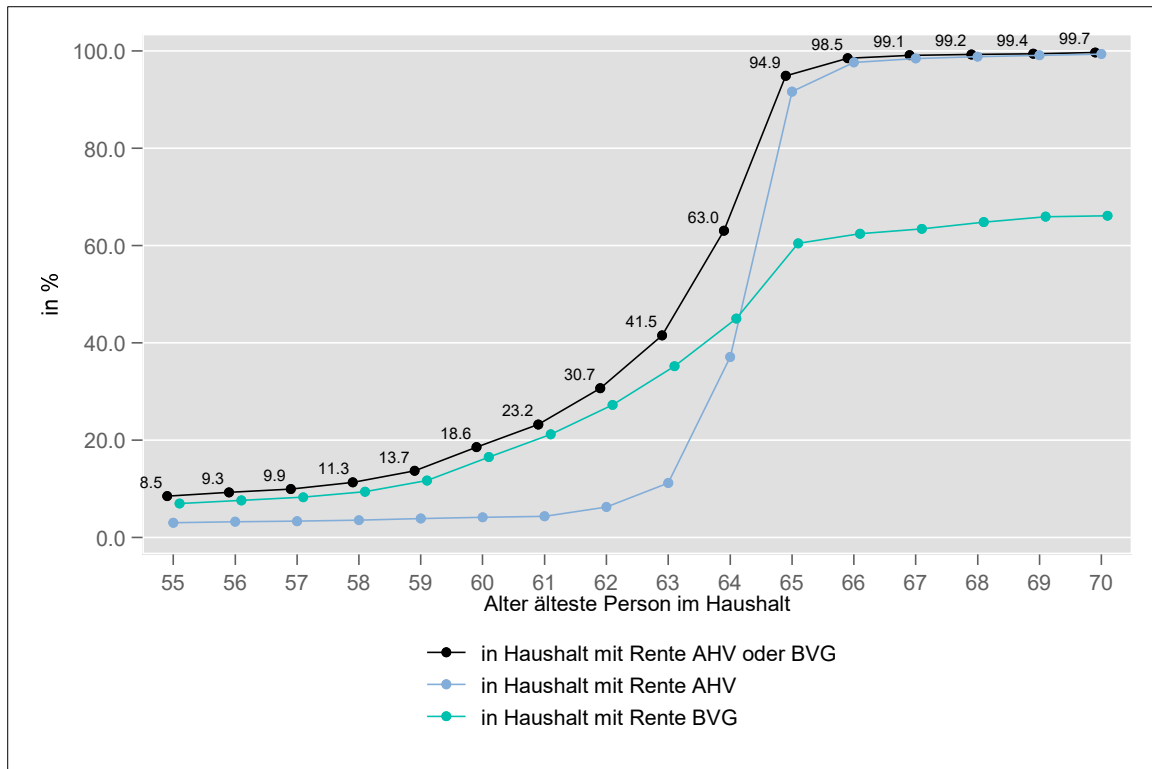
Ausschluss von Haushalten mit Altersrentnerinnen und -rentnern

Wie erwähnt ist der Indikator auf **Haushalte von Personen im erwerbsfähigen Alter** beschränkt. Diese konzeptuelle Einschränkung und der damit einhergehende Ausschluss von Haushalten mit Altersrentnerinnen und -rentnern erfolgt deshalb, da mit den verknüpften Administrativ- und Befragungsdaten keine verlässlichen Angaben zu den BVG-Renten verfügbar sind, welche für einen beträchtlichen Teil dieser Bevölkerungsgruppe eine substantielle Einkommensquelle darstellen. Auch ein allfälliger (teilweiser) Kapitalbezug des Altersguthabens kann mit den verfügbaren Daten nicht eruiert werden.

Konkret werden auch gemischte Haushalte mit Personen im Rentenalter und Personen vor Rentenalter ausgeschlossen (Mehrgenerationenhaushalte, Paare mit beträchtlichem Altersunterschied). Das Fehlen von Angaben zu Leistungen aus der 2. und 3. Säule (zusätzlich auch allfällige ausländische Renten) in den verknüpften Administrativ- und Befragungsdaten hat dadurch keine Verzerrung der Ergebnisse zur Folge.

Eine klare Altersgrenze zur Identifikation von Altersrentnerinnen und -rentnern existiert jedoch nicht. Das ordentliche Rentenalter beträgt im neusten verfügbaren WiSiER-Datenjahr 2015 65 Jahre bei Männern, respektive 64 Jahre bei Frauen. Ein allfälliger BVG-Rentenvorbezug ist jedoch bereits ab 58 Jahren, ein Vorbezug der AHV-Rente frühestens 2 Jahre vor Erreichen des ordentlichen Rentenalters möglich (zu Personen im Rentenübergang vgl. Braun-Dubler, Kaderli et al. 2022). Angaben zu Einkommen aus einem allfälligen BVG-Rentenvorbezug sind nur in den Steuerdaten enthalten und deshalb in der zukünftig verfügbaren Datengrundlage ohne Steuerinformationen nicht mehr vorhanden. Es wird also ohne Steuerdaten nicht mehr möglich sein, entsprechende Personen mit BVG-Rentenbezug vor Erreichen des ordentlichen Rentenalters individuell zu identifizieren, die ohne diese Angaben mit hoher Wahrscheinlichkeit irrtümlich als einkommensschwach kategorisiert würden. Aus diesem Grund ist eine angemessene Altersobergrenze zu definieren, um solche unerkannten Rentner/innen-Haushalte von den Analysen möglichst auszuschliessen.

Abbildung 2: Anteil Personen in Haushalten mit Rentenbezug AHV und BVG, nach Alter der ältesten Person



Quelle: WiSiER. Berechnungen BASS

Die genaue Festlegung erfolgt basierend auf empirischen Ergebnissen zum Rentenvorbezug mittels der WiSiER-Daten. **Abbildung 2** zeigt dazu den Anteil an Personen in Haushalten mit Rentenbezug AHV oder BVG nach Alter des ältesten Haushaltsmitglieds. Ersichtlich ist ein «Grundsockel» an Rentenbezügen bereits bei jüngeren Altersjahren vor Erreichen des minimalen Rentenvorbezugsalter von 58 Jahren. Es handelt sich dabei um Haushalte mit einer BVG Invalidenrente (welche in den Daten nicht von einer BVG Altersrente zu unterscheiden sind), sowie Haushalte mit einer AHV-Hinterlassenen-Rente (Witwen, Witwer, Waisen). Ab 59 Altersjahren ist dann eine verstärkte Zunahme des Bezugs von BVG-Renten festzustellen, ab 63 Altersjahren auch von AHV-Renten.

Für die Plausibilisierung haben wir angesichts dieser Ergebnisse den Schnitt bei Haushalten mit Personen über 60 Jahren vorgenommen, da ab Alter des ältesten Haushaltsmitglieds von 61 Jahren der Anteil an Personen in Haushalten mit BVG-Rente 20 Prozent überschreitet. Diese Grenzziehung ist ein pragmatischer Entscheid. Denkbar wäre es, diese Grenze auf 58 Jahre zu reduzieren, um potentielle BVG-Vorbeziehende vollständig auszuschliessen, oder aber höher anzusetzen, bspw. bei 63 Jahren. Über ein Drittel dieser Personen bei dieser höheren Grenze lebt jedoch bereits in einem Haushalt mit einer BVG-Rente und würde somit mit hoher Wahrscheinlichkeit irrtümlich als einkommensschwach kategorisiert, da im verknüpften Datensatz diese Einkommensquelle nicht mehr vorhanden ist (und BVG-Vorbeziehende in aller Regel finanziell bessergestellt sind). Diese Personen nicht auszuschliessen hätte nicht nur negative Auswirkungen auf die Qualität der Gesamtergebnisse, faktisch sind spezifische Analysen für diese Altersgruppe über 60 Jahre ohne Berücksichtigung dieser potentiellen wichtigen zusätzlichen Einkommensquelle auch wenig sinnvoll, auch weil der Fokus auf der Erwerbsphase liegt.

Tabelle 2 zeigt zusammenfassend die Grundgesamtheit sowie die einzelnen Selektionsschritte, welche für die Bestimmung der verwendeten Beobachtungen aus dem WiSiER-Datensatz für die Plausibilisierung vorgenommen wurden.

Tabelle 2: Grundgesamtheit und Selektionsschritte für die Plausibilisierung, WiSiER 2015

Grundgesamtheit und Selektionsschritte	Anzahl Personen	Anteil Ausgangsgesamtheit
Ständige Wohnbevölkerung in Privathaushalten WiSiER Grundgesamtheit (11 Kantone), Jahr 2015	4'271'274	100.0%
- abzüglich unplausibler oder fehlender Werte: Haushalte mit negativen Haushaltseinkommen (WiSiER, IK) oder fehlenden Haushaltseinkommen (WiSiER)	4'264'495	99.8%
- abzüglich Haushalte mit ausländischen Personen Aufenthaltsstatus B sowie weitere quellensteuerpflichtige Personen	3'716'807	87.0%
- abzüglich Haushalte von Personen unter 25 Jahren	3'679'239	86.1%
- abzüglich Haushalte mit potentiellen Altersrentner/innen (Personen über 60 Jahre)	2'489'246	58.3%

Quelle: WiSiER. Berechnungen BASS

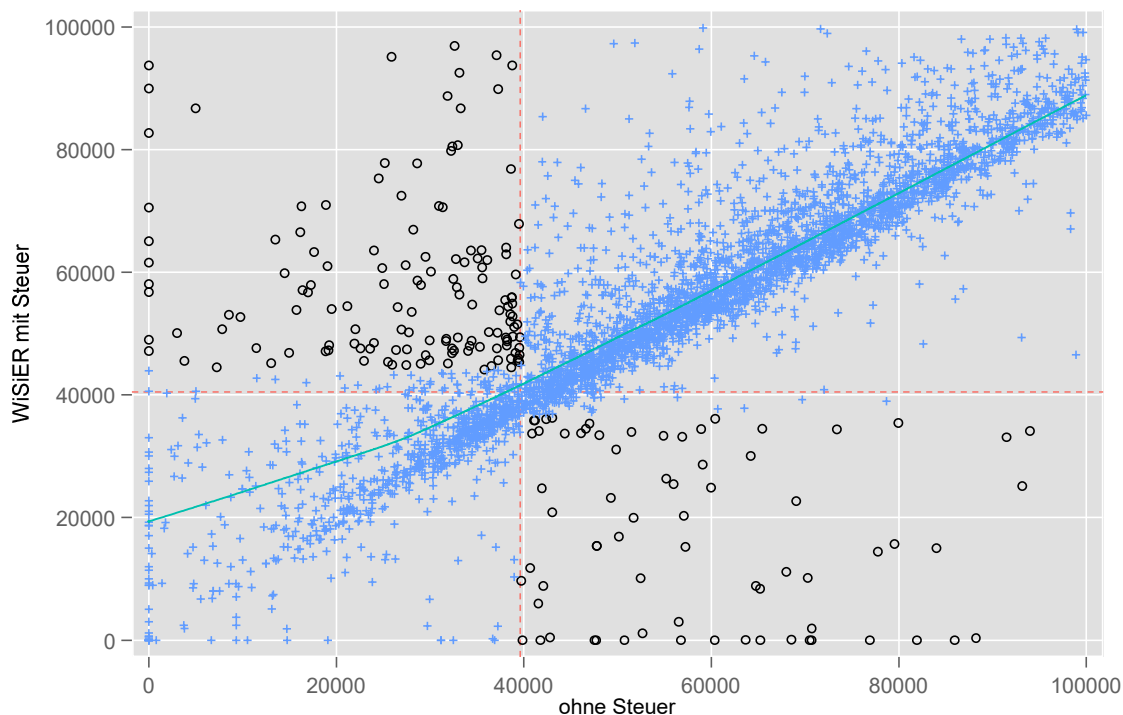
4.2 Ergebnisse der Plausibilisierung

In diesem Abschnitt werden die zentralen Befunde der Plausibilisierung präsentiert und diskutiert. Konkret werden im Folgenden für die zentrale Kategorisierung Haushaltstyp, im Anschluss auch für weitere Kategorisierungen wie bspw. Alter, Erwerbstätigkeit und Sozialhilfebezug jeweils die Plausibilisierungsergebnisse für den primären Indikator Erwerbseinkommensschwäche (P20) wie auch den alternativen Indikators «60% Median» auf Basis der verknüpften Daten ohne Steuer – in verschiedenen Varianten – ausgewiesen.

Dazu wird jeweils der Anteil der «korrekt» klassifizierten Personen ausgewiesen, wobei **«korrekt» klassifiziert verstanden wird als «gleich wie mit der Referenzdatenbasis WiSiER mit Steuer»**. Das Streudiagramm in **Abbildung 3** illustriert das Vorgehen, es bildet quasi das empirische Pendant zur Konfusions- oder Wahrheitsmatrix (vgl. Textkasten unten). Die Abbildung stellt das berechnete Haushalts-Äquivalenzeinkommen eines Individuums basierend auf WiSiER mit Steuer (y-Achse) dem berechneten Haushalts-Äquivalenzeinkommen eines Individuums basierend auf den verknüpften Daten ohne Steuer (x-Achse) gegenüber. Die roten gestrichelten Linien markieren jeweils die Grenzen des Indikators P20: Beobachtungen unterhalb dieser Linie werden basierend auf der jeweiligen Datengrundlage als «einkommensschwach» klassifiziert. Ist die Klassifizierung basierend auf beiden Datengrundlagen deckungsgleich, so befindet sich die Beobachtung entweder im unteren linken Quadranten oder im oberen rechten Quadranten, was mit einem Anteil von 90.4 Prozent für die überwiegende Mehrheit der Beobachtungen zutrifft. Eine fehlerhafte Klassifizierung liegt bei Beobachtungen im unteren rechten sowie im oberen linken Quadranten vor.

Bei der Beurteilung der Klassifikationsgüte lassen sich auch weitere Kenngrössen ableiten, etwa die Spezifität und Sensitivität. Die Sensitivität der Klassifikation ist dabei der Anteil der korrekt als einkommensschwach klassifizierten Personen, die Spezifität der Anteil der korrekt als nicht einkommensschwach klassifizierten Personen. Die Sensitivität beträgt hier 75.9 Prozent, die Spezifität 94.0 Prozent (siehe dazu **Tabelle 3** im Anhang mit der Konfusionsmatrix und abgeleiteten Kenngrössen).

Abbildung 3: Äquivalenz-Haushaltseinkommen WiSiER mit Steuer vs. verknüpfte Daten ohne Steuer mit Klassifikation «fuzzy»



Rot gestrichelte Linie: P20-Perzentilsgrenze. Korrekt klassifizierte Personen («fuzzy»-Klassifikation) sind mit blauem Kreuz, fehlerhafte Klassifikationen mit schwarzem Kreis ausgewiesen. Abbildung basiert auf einer Stichprobe zur besseren Darstellung.

Quelle: WiSiER. Berechnungen BASS

In Abbildung 3 ist die Klassifikationsgüte gemäss dem **«fuzzy»-Kriterium** ersichtlich: Beobachtungen, die nur knapp falsch klassifiziert wurden, d.h. innerhalb eines Grenzbereichs von ± 5 Perzentilen um den Grenzwert P20 gemäss Referenzdatenbasis WiSiER mit Steuer liegen, wird eine Klassifikation in beide Gruppen durch die zu plausibilisierende Datenbasis als akzeptabel und folglich korrekt angesehen. Diese Beobachtungen sind in der Abbildung mit einem blauen Kreuz ausgewiesen. Der entsprechende Anteil steigt gegenüber dem «sharp»-Kriterium um einige Prozentpunkt auf 93.9 Prozent. Die auch mit dem «fuzzy»-Kriterium fehlerhaft klassifizierte Beobachtungen sind mit einem runden schwarzen Kreis ausgewiesen.

Bestimmung der Güte einer Klassifikation mittels Konfusionsmatrix

Die Qualität einer binären Klassifikation lässt sich mittels einer sog. **Konfusions- oder Wahrheitsmatrix** abbilden. Dabei werden die Ergebnisse der auf Grundlage der beiden Datensätze (WiSiER mit Steuer vs. verknüpfte Daten ohne Steuer) erfolgten Zuweisungen in Form einer Kreuztabelle dargestellt: korrekte positive und negative Zuweisungen, falsch negative Zuweisungen (α -Fehler) und falsch positive Zuweisungen (β -Fehler). Aus einer solchen Matrix lassen sich Kenngrössen wie insbesondere der in den nachfolgenden Abbildungen jeweils ausgewiesene Anteil **korrekt klassifizierter Beobachtungen** ableiten.¹³ «Korrekt klassifiziert» heisst dabei präzise ausgedrückt «gleich klassifiziert wie basierend auf WiSiER (mit Steuerdaten)», wobei für die Interpretation der Ergebnisse zu beachten ist, dass auch die Zuweisung durch WiSiER mit Steuer

¹³ Bei der Beurteilung der Klassifikationsgüte lassen sich auch weitere Kenngrössen ableiten, so etwa die Spezifität und Sensitivität. Die Konfusionsmatrix sowie diese Kenngrössen für die P20-Klassifikation durch die verknüpften Daten ohne Steuer gegenüber der Referenzdatenbasis WiSiER sind in Tabelle 3Tabelle 3 im Anhang ausgewiesen. Ansonsten verzichten wir in der Plausibilisierung auf die Präsentation dieser sekundären Informationen.

fehlerhaft sein kann.

Fuzzy-Klassifikation: Mit Blick auf die Eignungsbewertung wird ergänzend auch graduell zwischen wesentlich fehlerhaften Zuweisungen und geringfügig fehlerhaften Zuweisung (knapp unter/über der relevanten Perzentilgrenze) differenziert, in Anlehnung an das Konzept der «fuzzy confusion matrix». Die Überlegung dahinter ist, dass solche geringfügig fehlerhaften Zuweisungen aus einer inhaltlichen Sichtweise weniger problematisch sind, weil sich die Betroffenen knapp unter oder über dem Grenzwert in einer relativ ähnlichen Situation befinden. Als Kriterium für eine tolerierbare geringfügige Abweichung wird eine **Fehlreihung um maximal +/- 5 Perzentile** verwendet. Konkret würden somit mit der fuzzy-Klassifikation bei der Armutskennzahl P20 «false positives» bis zum 25 Perzentil und «false negatives» bis zum 15 Perzentil (jeweils bei der Klassifikation mittels WiSiER mit Steuer) als immer noch «korrekt» ausgewiesen.¹⁴

Neben der Frage nach der korrekten Zuteilung einzelner Beobachtungen wird im Folgenden jeweils auch die Abweichungen der berechneten Anteilswerte des Indikators für einzelne Kategorien wie bspw. Haushaltstypen ausgewiesen. Dies ist relevant mit Blick auf die Identifikation eines allfälligen Bias, einer systematischen Verzerrung. Konkret manifestiert sich eine Verzerrung auf der Aggregatenebene in Form von Differenzen bei berechneten Parametern wie Anteils- und Mittelwerten. Bei der nachfolgenden Plausibilisierung liegt der Fokus auf **Abweichungen bei den Anteilswerten einkommensschwacher Personen bei den einzelnen Bevölkerungsgruppen**. Neben dem Ausmass der Abweichung gegenüber dem Referenzwert WiSiER mit Steuer ist auch relevant, inwiefern die in den Referenzdaten WiSiER mit Steuer beobachteten Muster erhalten bleiben (bspw. Rangfolge der Analysekatoren). Dies ist in den Abbildungen jeweils visuell ersichtlich.

4.2.1 Ergebnisse für die zentrale Gruppierungsvariable Haushaltstyp

In diesem Abschnitt werden die Ergebnisse für die Gruppierungsvariable Haushaltstyp nach Alter jüngstes Kind ausgeführt und diskutiert – dies auch exemplarisch und als Lese- und Interpretationshilfe für alle anderen entsprechenden Abbildungen hier und im separaten Auswertungsdokument. Beim Haushaltstyp (nach Alter jüngstes Kind) handelt es sich um eine für das Armutsmonitoring zentrale Differenzierung, welche auch die zentralen Herausforderungen aufzeigt, die sich mit der neuen Datenbasis als Folge der fehlenden Einkommenskomponenten stellen.

Dabei werden zuerst die Ergebnisse für den Indikator P20 basierend auf der Standardvariante «verknüpfte Daten ohne Steuer» präsentiert, danach folgen die Ergebnisse für den Indikator P20 basierend auf den optimierten verknüpften Daten ohne Steuer, dann die Ergebnisse für den alternativen Indikator «60% Median» und schliesslich die Ergebnisse für den Indikator P20 basierend auf der Variante «nur Erwerbseinkommen».

Ergebnisse für den Indikator P20 mit Datenbasis verknüpfte Daten ohne Steuer

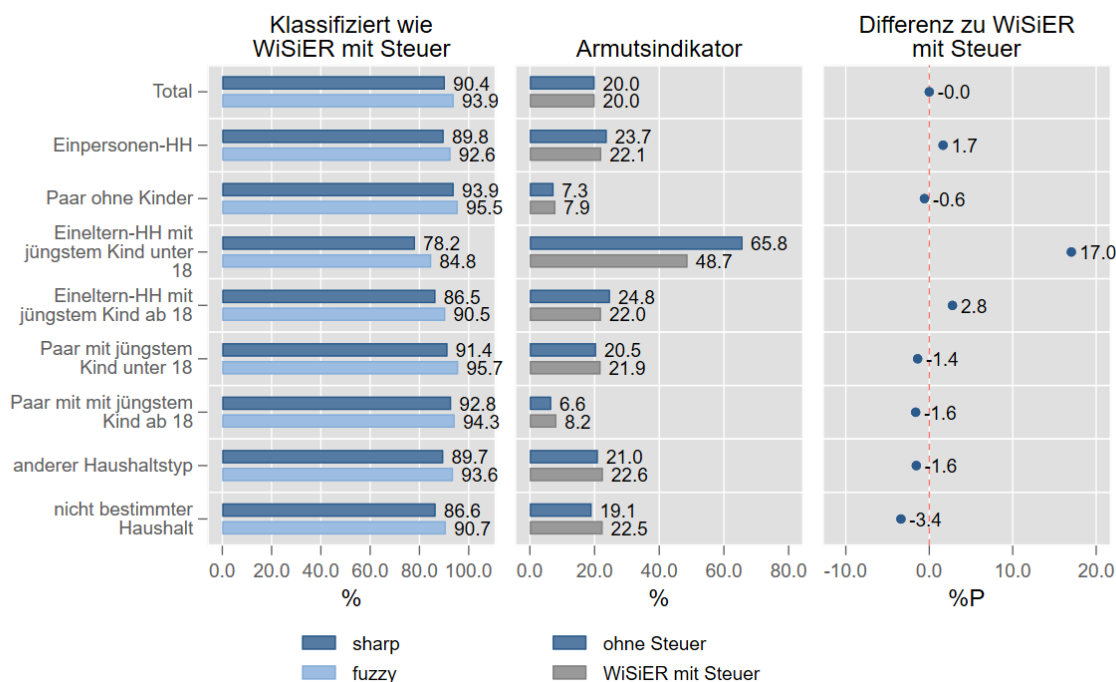
Die Plausibilisierungsergebnisse für den primären Indikator P20 berechnet auf Basis der verknüpften Daten ohne Steuer für die Gruppierungsvariable Haushaltstyp (nach Alter jüngstes Kind) zeigt **Abbildung 4**. In der Grafik aussen links ist die **Klassifikationsgüte** ersichtlich, die oben bereits auf Ebene

¹⁴ Bei der Armutsgefährdungsgrenze ist der Grenzwert 60% des Medians des Haushaltseinkommens. Als bei der «fuzzy»-Klassifikation verwendete akzeptable Bandbreite bei der Referenzdatenbasis WiSiER wird entsprechend 55%, respektive 65% des Medians verwendet.

des Gesamttotals ausgeführt wurde. Es wird hier der Anteil Personen ausgewiesen, die bezüglich des Indikators «korrekt» klassifiziert wurden. Hier und im Folgenden wird «korrekt» klassifiziert jeweils als «gleich wie mit der Referenzdatenbasis WiSiER mit Steuer» verstanden. Dies ist die zentrale Kenngrösse, welche sich aus einer sogenannten Konfusionsmatrix ableiten lässt (vgl. Textkasten oben).

Ergänzend wird in der Grafik aussen links neben dem Ergebnis der «sharp»-Klassifikation (dunkelblau) auch das Ergebnis der «fuzzy»-Klassifikation (hellblau) ausgewiesen. Letztere geht mit durch die Referenzdatenbasis WiSiER mit Steuer «knapp» klassifizierten Personen gesondert um, da bei diesen Grenzfällen eine falsche Zuteilung weniger problematisch ist. «Knapp» klassifiziert bedeutet dabei innerhalb der Bandbreite vom 15. bis 25. Perzentil (+/- 5 Perzentile um den Grenzwert P20) – bei diesen Personen wird eine Klassifikation in beide Gruppen durch die zu plausibilisierende Datenbasis als akzeptabel und folglich korrekt angesehen.

Abbildung 4: Plausibilisierung P20 - Verknüpfte Daten ohne Steuer, nach Haushaltstyp mit Alter jüngstes Kind



Lesebeispiel Spalte links: Total sind 90.4 Prozent aller Personen auf Grundlage der verknüpften Daten ohne Steuer (gegenüber der Referenzdatenbasis WiSiER mit Steuer) «korrekt» klassifiziert, 93.9 Prozent sind es bei Anwendung des «fuzzy»-Kriteriums. Von allen Haushaltstypen am schlechtesten gelingt die Klassifikation bei Personen in Eineltern-Haushalten mit jüngstem Kind unter 18 Jahren (78.2% «sharp», 84.8% «fuzzy»).

Lesebeispiel mittlere und rechte Spalte: Der Anteilswert des Indikators P20 wird bei Eineltern-Haushalten mit jüngstem Kind unter 18 Jahren mit den verknüpften Daten ohne Steuer um 17.0 Prozentpunkte überschätzt (Anteil von 65.8% mit den verknüpften Daten ohne Steuer gegenüber 48.7 Prozent mit der Referenzdatenbasis WiSiER mit Steuer).

Quelle: WiSiER. Berechnungen BASS

Wie in der Grafik aussen links ersichtlich ist, sind im Total 90.4 Prozent aller Personen auf Grundlage der verknüpften Daten gemäss der «sharp»-Klassifikation korrekt klassifiziert. Von allen Haushaltstypen am schlechtesten gelingt die Klassifikation mittels verknüpfter Daten bei Personen in Eineltern-HH mit jüngstem Kind unter 18 Jahren, wo mit lediglich 78.2 Prozent deutlich weniger Personen korrekt klassifiziert wurden. Mit der «fuzzy»-Klassifikation erhöht sich dieser Wert auf 84.8 Prozent, er bleibt aber auch hier unterdurchschnittlich. Das Vorgehen bei der Bestimmung der Klassifikationsgüte, konkret der «fuzzy»-Klassifikation, wird in **Abbildung 14** im Anhang mittels Streudiagrammen illustriert.

Für die Plausibilisierung ist neben der Klassifikationsgüte auf der individuellen Ebene auch die Identifikation eines allfälligen Bias, der systematischen (nicht-zufälligen) Verzerrung der Schätzer, von Interesse, wie sie sich in Form von **abweichenden Anteilswerten** bezüglich des Indikators manifestiert. Diese Anteilswerte an einkommensschwachen Personen (P20) für die einzelnen Haushaltstypen sind in der mittleren Grafik von Abbildung 4 ersichtlich – jeweils für die zu plausibilisierende Datenbasis, hier die verknüpften Daten ohne Steuer, wie auch für die Referenzdatenbasis WiSiER mit Steuer. Da bei diesen Anteilswerten auch bereits absolut gesehen geringfügige Unterschiede substantiell relevant sein können, ist in der Grafik aussen rechts schliesslich die daraus abgeleitete Differenz in den Anteilswerten in Prozentpunkten ausgewiesen. **Positive Werte zeigen hier eine Überschätzung des Anteils einkommensschwacher Personen, negative Werte eine Unterschätzung an** (gegenüber der Referenzdatenbasis WiSiER mit Steuer). Eine Überschätzung der Einkommensschwäche bei einer bestimmten Kategorie ist dabei ein Hinweis darauf, dass bei der entsprechenden Bevölkerungsgruppe weitere Einkommensquellen relevant sind, die wohl in WiSiER mit Steuer, nicht aber in der neuen Datenbasis vorhanden sind (vgl. die unten erwähnten Eineltern-Haushalte). Eine Unterschätzung bei einer Kategorie hängt meist nicht direkt mit dieser zusammen, sondern ist die indirekte Folge davon, dass die Einkommensschwäche von anderen Kategorien überschätzt wird. Da der Anteil im Total beim Indikator P20 per Definition immer 20 Prozent beträgt, sind lediglich auf Ebene der einzelnen Kategorien Differenzen möglich. In Abbildung 4 zeigt sich diesbezüglich, dass das grundsätzliche Muster bei den zwei Datengrundlagen bei den einzelnen Kategorien durchaus vergleichbar ist – Haushalte mit hohem, mittlerem und jene mit tiefem Armutsrisiko bleiben dieselben. Allerdings finden sich kleinere und grössere substantielle Abweichungen bei einzelnen Haushaltstypen.

Besonders ausgeprägte Unterschiede zeigen sich bei Eineltern-HH mit jüngstem Kind unter 18 Jahren, bei welchen mit den verknüpften Daten ohne Steuer der Anteil der P20 um 17.0 Prozentpunkte überschätzt wird (65.8% mit den verknüpften Daten ohne Steuer gegenüber 48.7 Prozent mit der Referenzdatenbasis WiSiER mit Steuer). Der Grund für diese wesentliche Verzerrung liegt primär darin, dass Unterhaltszahlungen durch den anderen Elternteil für viele dieser Haushalte eine substantielle Einkommensquelle darstellen, und diese wohl in den Steuerdaten von WiSiER, nicht aber in den verknüpften Daten ohne Steuer vorhanden sind (vgl. dazu auch zur Illustration das Streudiagramm in **Abbildung 16** im Anhang).¹⁵

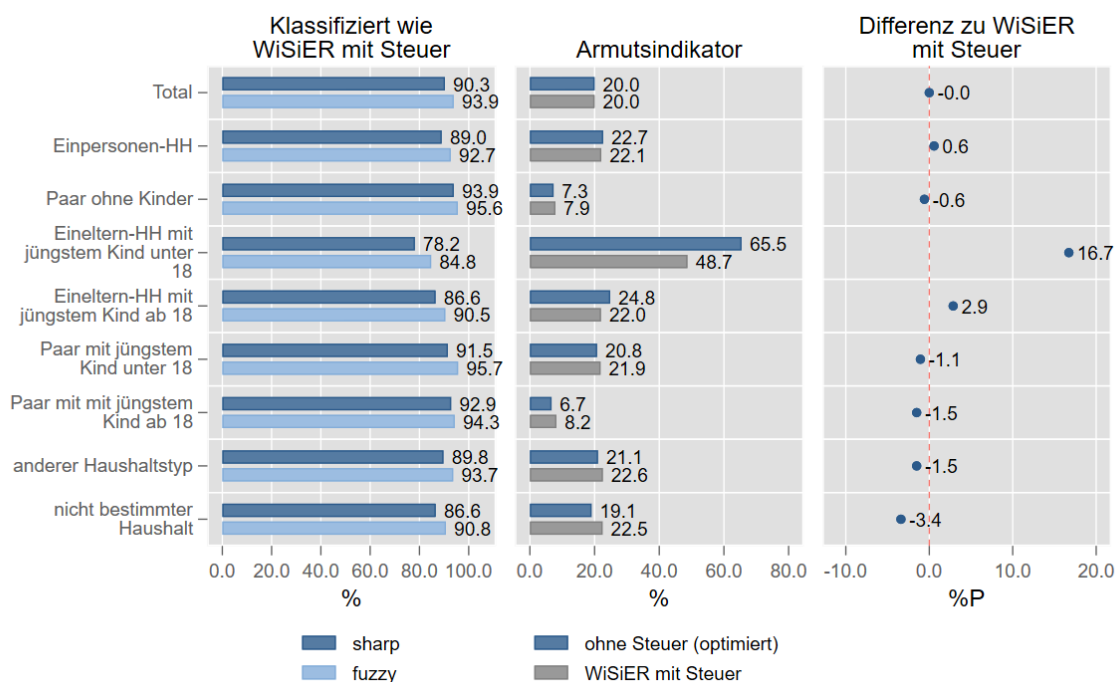
Die folgenden zahlreichen weiteren Abbildungen für weitere Varianten bei der zu plausibilisierenden Datenbasis wie auch beim Indikator sind identisch aufgebaut und sind analog zu interpretieren, wie hier exemplarisch ausgeführt.

¹⁵ Es sei darauf hingewiesen, dass nur Unterhaltszahlungen an minderjährige Kinder steuerlich abzugsfähig (und entsprechend auch bei der empfangenden Person auszuweisen) sind und deshalb nur für diese Informationen in den Steuerdaten vorliegen. Bei erwachsenen Kindern in Ausbildung besteht auch bei WiSiER ein grösserer blinder Fleck bezüglich allfälliger Unterhaltszahlungen.

Vergleich mit den Ergebnissen für den Indikator P20 mit der Datenbasis der optimierten verknüpften Daten ohne Steuer

Mit den optimierten verknüpften Daten ohne Steuer – konkret der Modellierung des Nettolohns und der Kinderzulagen – verändert sich das Bild nur geringfügig, wie **Abbildung 5** zeigt. Der Anteil der korrekt klassifizierten Personen verändert sich nicht systematisch, es tritt bezüglich der Klassifikationsgüte keine Verbesserung ein.

Abbildung 5: Plausibilisierung P20 - Verknüpfte Daten ohne Steuer (**optimiert**), nach Haushaltstyp mit Alter jüngstes Kind



Quelle: WiSiER. Berechnungen BASS

Leicht anders sieht es bei den Verzerrungen auf der Aggregatebene aus, d.h. bei den Abweichungen bei den Anteilswerten: Die Differenzen in den Anteilswerten nach Haushaltstyp (mit Alter jüngstes Kind) verringern sich gegenüber den nicht optimierten verknüpften Daten ohne Steuer (Abbildung 4 oben) marginal. Die Differenz von +17 Prozentpunkten beim Anteil einkommensschwacher Personen in Eineltern-HH mit jüngstem Kind unter 18 sinkt bspw. auf +16.7 Prozentpunkte. Obwohl also eine Verbesserung stattfindet, ist diese gering, und die beobachteten grösseren und potentiell problematischen Verzerrungen, respektive Unterschiede bei den Anteilswerten, bleiben davon grösstenteils unbeeinflusst und weiterhin bestehen.

Wie wirkt sich die vorgenommene Optimierung der Datengrundlage konkret aus, wie sie oben beim Vorgehen unter der entsprechenden Berechnungsvariante ausgeführt wurde? Haushalte mit Erwerbseinkommen werden mit der Berechnung des Nettolohns tendenziell präziser (und faktisch relativ gegenüber anderen Haushalten ärmer) abgebildet. Als Konsequenz werden die Differenzen in den Anteilswerten etwa bei IV-Rentenbeziehenden vs. regulären Haushalten verringert (vgl. die entsprechenden Abbildungen im separaten Auswertungsdokument). Die Kinderzulagen wiederum erhöhen die Präzision bei Haushalten mit Kindern und schlagen etwa bei Eineltern-HH mit minderjährigen Kindern zu Buche (vgl. die entsprechende Veränderung der Differenz zwischen Abbildung 4 und Abbildung 5 um -0.3 Prozentpunkte). Der Effekt dieser Optimierung ist gering, da die Beträge bei den

4 Plausibilisierung der verknüpften Daten ohne Steuerinformationen zur Bestimmung der einkommensschwachen Bevölkerung

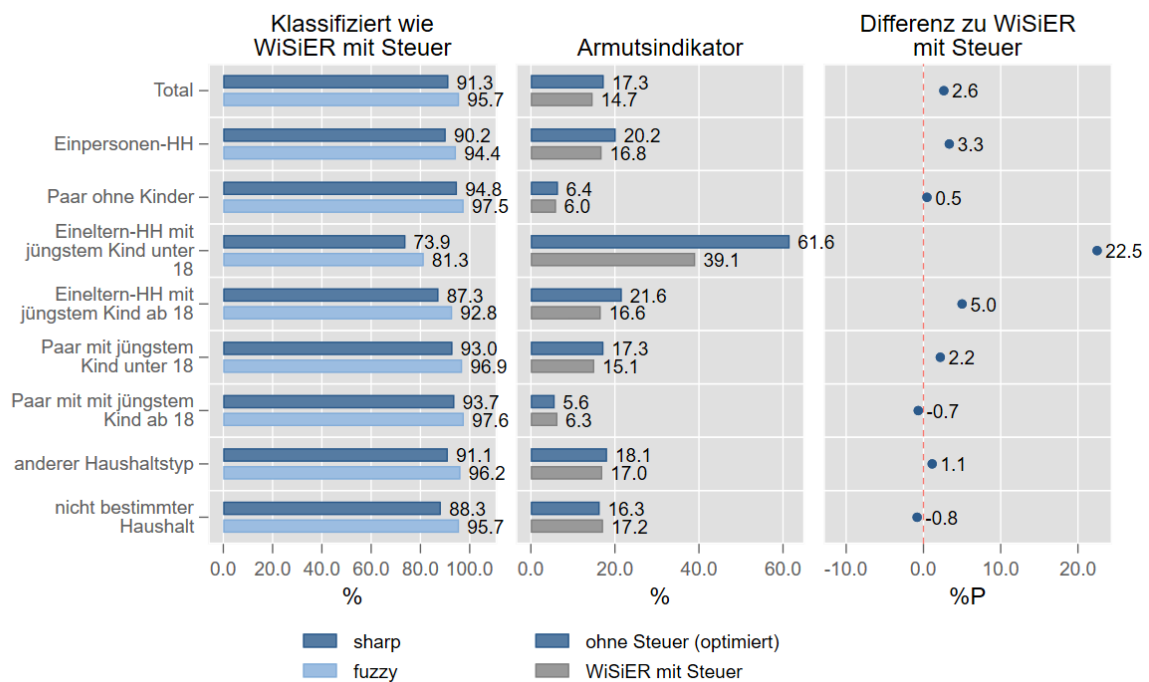
Kinderzulagen gering sind und zudem auch noch an Haushalte mit mehreren Personen (Kind und Elternteile) fliessen, auf welche dieser Betrag dann bei der Berechnung des Haushaltsäquivalenzeinkommens «verteilt» wird.

Das kaum oder nur geringe Verbesserungen bei den Ergebnissen als Folge dieser Optimierung sichtbar sind, ist kein grundsätzliches Argument gegen solche Optimierungen, falls diese wie hier mit geringem Aufwand vorgenommen werden können. Aus diesem Grund präsentieren wir im Folgenden auch jeweils primär diese Ergebnisse auf Basis der optimierten verknüpften Daten ohne Steuer. Das beschränkte Optimierungspotenzial verdeutlicht zugleich, dass die Folgen der grossen Lücken bei den Einkommenskomponenten nicht einfach mit der Modellierung oder der zusätzlichen Integration in den verknüpften Datensatz von Informationen zu sekundären Einkommenskomponenten, wie etwa den Kinderzulagen, kompensiert werden können. Relevant sind letztlich primär jene Einkommenskomponenten, die substantiell zum Haushaltseinkommen beitragen, wie etwa Alimente oder BVG-Rentenzahlungen.

Vergleich mit den Ergebnissen für den alternativen Indikator «60% Median» (Datenbasis verknüpfte Daten ohne Steuer optimiert)

Die Plausibilisierungsergebnisse für den alternativen Indikator «60% des Medians» (anstelle P20), für die Variable Haushaltstyp (mit Alter jüngstes Kind) und mit den optimierten verknüpften Daten ohne Steuer finden sich in **Abbildung 6**.¹⁶

Abbildung 6: Plausibilisierung **60% Median** - Verknüpfte Daten ohne Steuer (optimiert), nach Haushaltstyp mit Alter jüngstes Kind



Quelle: WiSiER. Berechnungen BASS

¹⁶ Gegenüber den Ergebnissen mit den nicht optimierten verknüpften Daten ohne Steuer sind bei den optimierten Daten die Differenzen bei den Anteilswerten etwas geringer, und der Anteil «korrekt» klassifizierter Personen ist marginal höher, vgl. die entsprechende Auswertung im separaten Dokument.

Der Anteil korrekt «sharp» klassifizierter Personen fällt gegenüber dem Indikator P20 mit 91.3 Prozent gegenüber 90.3 geringfügig besser aus (vgl. mit Abbildung 5 oben).

Die Verzerrungen bei den Anteilswerten sind bezüglich des generellen Musters vergleichbar mit dem Indikator P20, absolut gesehen fallen die Abweichungen allerdings höher aus. Dies liegt primär an der unterschiedlichen Berechnungsweise der beiden Indikatoren – P20 als gänzlich relatives, 60% Median als nur bedingt relatives Armutsmass: Bei der Armutsgefährdung fällt nicht wie bei P20 das Total per Definition gleich aus, sondern es besteht bereits hier eine nicht unwesentliche Differenz: Der Anteil Armutsgefährdeter bei WiSiER mit Steuer beträgt 14.7 Prozent, bei den verknüpften Daten ohne Steuer (optimiert) 17.3% – eine Überschätzung beim Total um +2.6 Prozentpunkten, die bei den einzelnen Unterkategorien zu den dortigen Abweichungen jeweils immer noch dazu kommt.

Vergleich mit den Ergebnissen für den Indikator P20 mit der Datenbasis «nur Erwerbseinkommen»

Die Plausibilisierung wurde auch für die alternative Datenbasis «nur Erwerbseinkommen»¹⁷ vorgenommen und sind in **Abbildung 7** ausgewiesen. Die Abweichungen gegenüber der Referenzbasis WiSiER-Haushaltseinkommen mit Steuer fallen dabei zwingend grösser aus. Dies deshalb, da bei diesem Indikator auf die wichtigste und zugleich präzise erfasste Einkommenskomponente fokussiert wird: der Lohn (sowie aufgrund des sachlichen Zusammenhangs zusätzlich allfällige ALV-Leistungen). Weitere Einkommenskomponenten, wie etwa bedarfsabhängige Sozialleistungen und Leistungen von anderen Sozialversicherungen (inkl. IV), werden bewusst weggelassen. Von Interesse ist bei dieser alternativen Datenbasis deshalb nicht die Frage, ob sie besser performt als die verknüpften Daten ohne Steuer, was gar nicht möglich ist, sondern wie gross bei jeweils welchen Gruppen die Abweichungen gegenüber der Datenbasis WiSiER mit Steuer aufgrund dieser engeren Konzeptualisierung ausfallen.

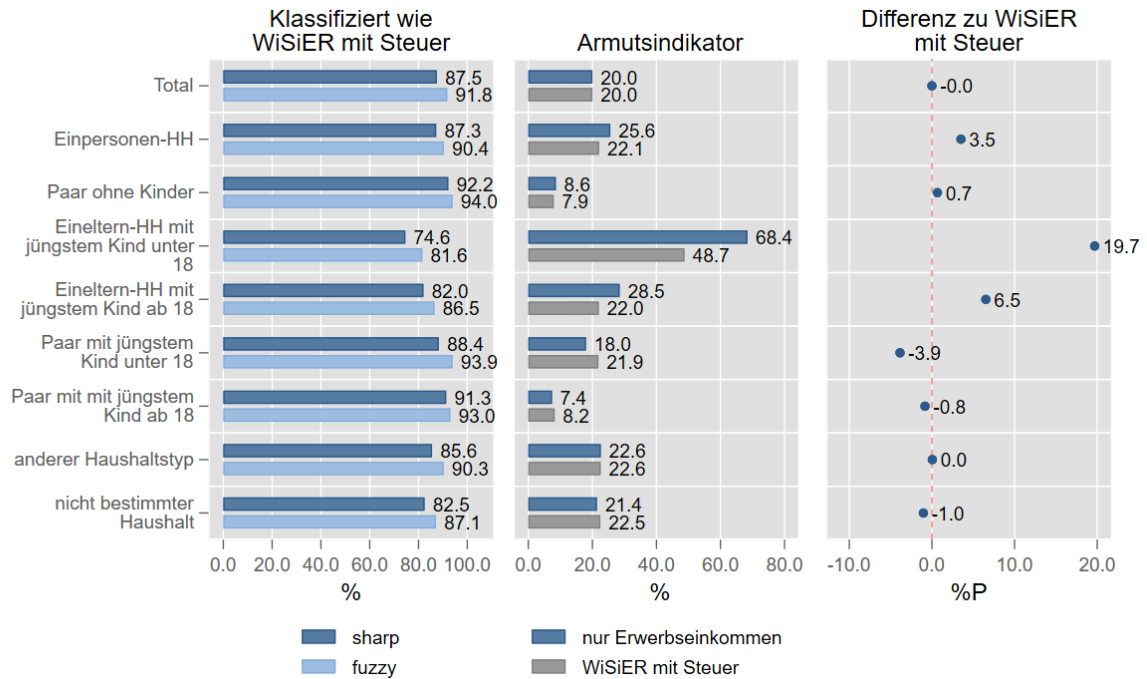
Wie Abbildung 7 zeigt, sind bei dieser Datenbasis «nur Erwerbseinkommen» wie erwartet grössere Differenzen in den Anteilswerten zur Referenzdatenbasis WiSiER mit Steuer sichtbar als bei den oben präsentierten Ergebnissen mit den vollständigeren Datenbasis-Varianten verknüpften Daten ohne Steuer. Das beobachtete allgemeine Muster über die Haushaltstypen hinweg ist jedoch auch hier dasselbe: Auch hier weisen Eineltern-HH mit jüngstem Kind unter 18 das höchste Risiko für Einkommensschwäche auf – die Überschätzung beim Anteil der einkommensschwachen Personen (P20) beträgt +19.7 Prozentpunkte, bei den optimierten verknüpften Daten ohne Steuer waren es +16.7 Prozentpunkte. Neu zeigt sich eine gewisse Unterschätzung bei den Paarhaushalten mit jüngstem Kind unter 18 Jahren von -3.9 Prozentpunkten.

Wie erwartet, sind die Abweichungen hier gegenüber WiSiER mit Steuerdaten nochmals etwas grösser als bei den verknüpften Daten ohne Steuer. Als Indikator für das Gesamteinkommen eines Haushalts und für Armut ist diese Datenbasis «nur Erwerbseinkommen» somit noch weniger als die anderen oben präsentierten Varianten geeignet. Hingegen ist der Vorteil dieser Datenbasis, dass sie konzeptuell relevant und zugleich empirisch präzise bestimmt ist: Wenn das Erwerbseinkommen eines Haushalts in einer Analyse im Vordergrund steht, dann ist dieser Indikator durchaus zweckmässig.

¹⁷ Auch hier wurde, wie bei den optimierten verknüpften Daten, das Nettoeinkommen berechnet.

4 Plausibilisierung der verknüpften Daten ohne Steuerinformationen zur Bestimmung der einkommensschwachen Bevölkerung

Abbildung 7: Plausibilisierung P20 – **nur Erwerbseinkommen**, nach Haushaltstyp mit Alter jüngstes Kind



Quelle: WiSiER. Berechnungen BASS

4.2.2 Ergebnisse zu weiteren Gruppierungsvariablen

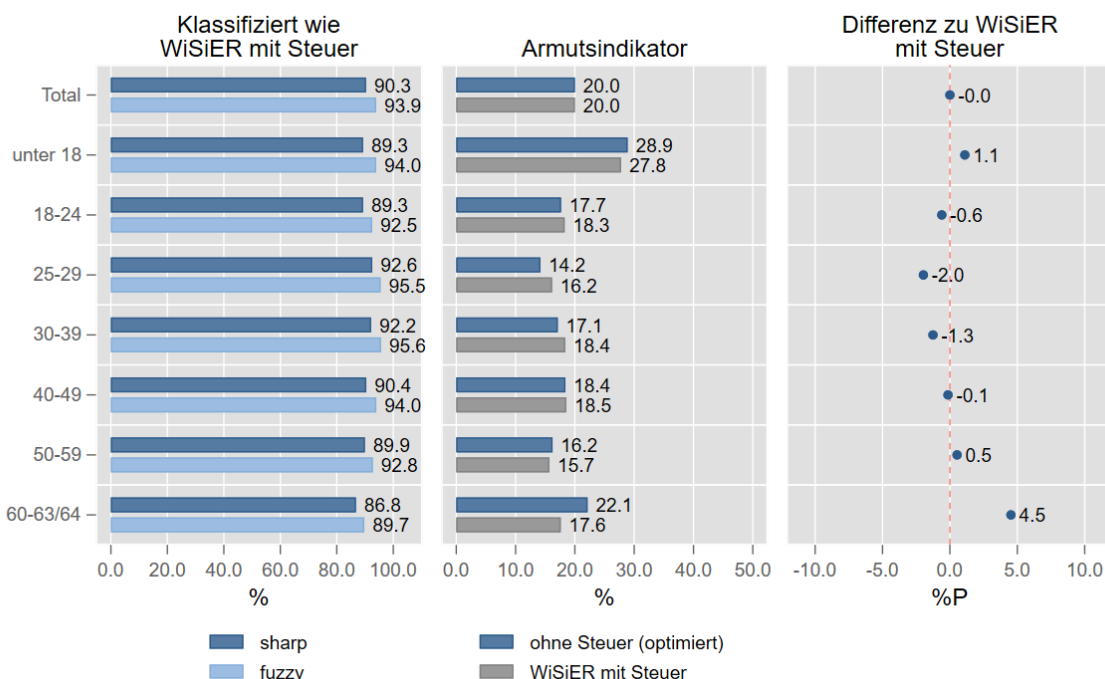
Im Folgenden werden relevante Ergebnisse für weitere Gruppierungsvariablen kurz diskutiert. Eine Auswahl der zentralen Abbildungen ist hier dargestellt, im separaten Auswertungsdokument finden sich auch alle weiteren Abbildungen.

Gruppierungsvariable Alter

Das allgemeine Muster bezüglich Rangfolge und Ausmass des Armutsrisikos, wie es mit WiSiER mit Steuer sich zeigt, bleibt für die Altersgruppen auch mit der Datenbasis ohne Steuer (optimiert) erhalten (**Abbildung 8**). Auch die Differenzen beim Anteil P20 zwischen den beiden Datenbasen sind nicht allzu stark, mit Ausnahme bei der Alterskategorie ab 60. Es ist zu beachten, dass aufgrund des vorgenommenen Ausschlusses von Haushalten von potentiellen Rentnerinnen und Rentnern in dieser Kategorie nur 60-Jährige und keine älteren Personen enthalten sind. Dennoch ist die Verzerrung in Form einer Überschätzung des Anteils einkommensschwacher Personen um +3.5 Prozentpunkte bereits bei diesen 60-jährigen Personen wesentlich – diese Altersgruppe wird so fälschlicherweise nach den Minderjährigen zur Gruppe mit dem zweithöchsten Risiko (22.1%) gezählt, obwohl ihr tatsächliches Risiko gemäss WiSiER-Datenbasis mit Steuer am tiefsten ist (17.6%). Zu beachten ist: Bei einer altersmässig höheren Grenzziehung bei den Rentnerhaushalten, wenn ältere Jahrgänge in der Kategorie ab 60 Jahren auch noch enthalten wären, würde sich diese Verzerrung noch verstärken.

Dieselbe Analyse nach Altersgruppen nach der ältesten Person **im Haushalt** (nicht wie hier der Zielperson) führt zu weitgehend identischen Ergebnissen (hier ohne Abbildung, vgl. separates Auswertungsdokument).

Abbildung 8: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Altersgruppe**



Quelle: WiSiER. Berechnungen BASS

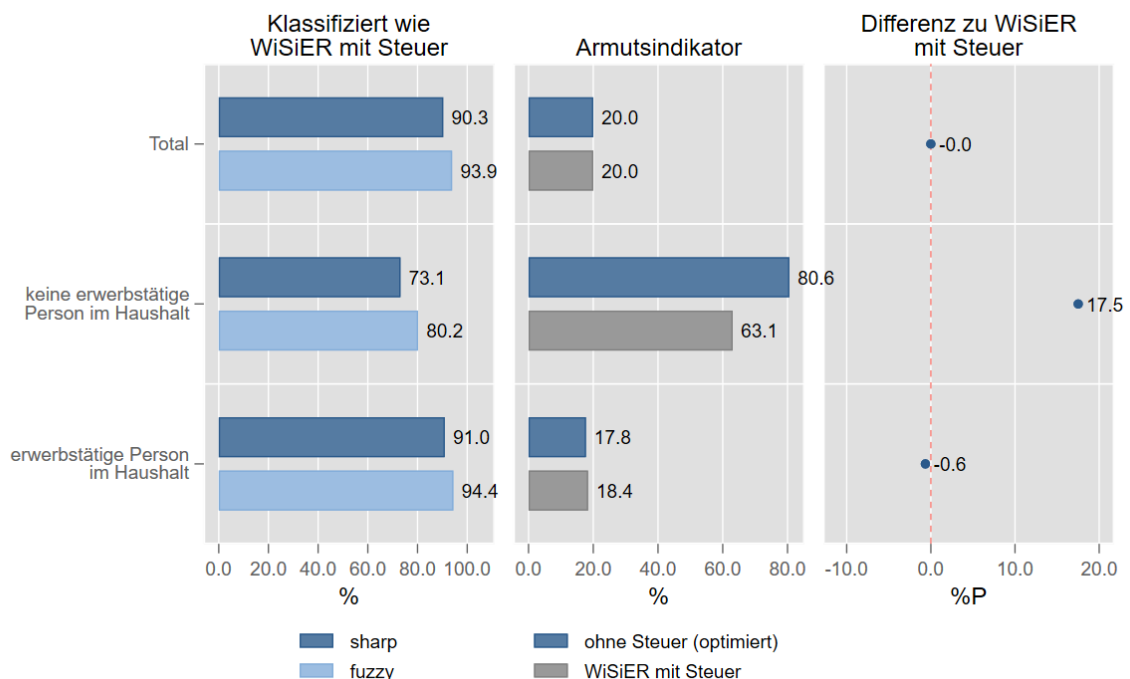
Gruppierungsvariablen Erwerbstätigkeit und Typ des Erwerbseinkommens

Bei der Differenzierung der Personen danach, ob sie in einem **Haushalt mit erwerbstätigen Personen** leben (**Abbildung 9**), zeigt sich die P20-Klassifikation bei Haushalten ohne Erwerbstätigkeit wesentlich weniger zuverlässig (73.1% korrekt «sharp» klassifiziert, 80.2 «fuzzy»). Es kommt zu einer wesentlichen systematischen Verzerrung: Einkommensschwache Personen in Haushalten ohne Erwerbstätigkeit werden um 17.5 Prozentpunkte überschätzt. Verantwortlich sind hierfür primär BVG-Renteneinkünfte, die häufig entscheidend dabei sind, einen Haushalt ohne erwerbstätige Personen aus der Einkommensschwäche zu heben, und die aufgrund der fehlenden Steuerinformationen bei den verknüpften Daten ohne Steuer nicht verfügbar sind.

Wird bei Haushalten nach der **Art des Haupterwerbseinkommens** eines Haushalts unterschieden (angestellt vs. AHV-selbständig), so zeigt sich bei den Anteilswerten P20 das grundsätzlich gleiche Muster, nämlich dass bei AHV-Selbständigkeit das Risiko rund doppelt so hoch ist. Dabei wird das Risiko bei AHV-Selbständigkeit jedoch um +2.7 Prozent höher eingestuft und somit überschätzt, (ohne Abbildung, vgl. separates Auswertungsdokument).

Im separaten Auswertungsdokument finden sich zusätzliche Auswertungen nach Erwerbstätigkeit und Art der Erwerbstätigkeit (angestellt vs. AHV-selbständig) auf Ebene Person, wobei sich diese Ergebnisse weitgehend mit den obigen auf der Haushaltsebene decken.

Abbildung 9: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Erwerbstätigkeit Haushalt**



Quelle: WiSiER. Berechnungen BASS

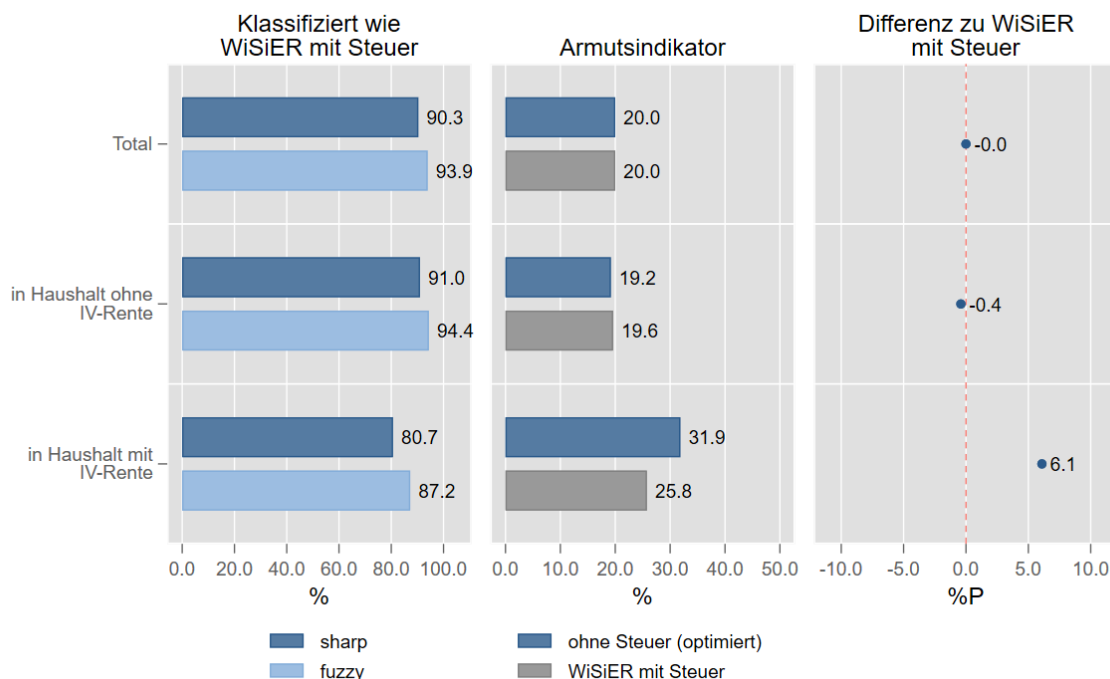
Haushalte mit IV- oder AHV-Rente sowie Ergänzungsleistungen

Wie oben bereits für andere Bevölkerungssegmente aufgezeigt (Altersgruppe ab 60, Haushalte ohne Erwerbstätigkeit), führt das Fehlen von BVG-Renten-Informationen auch bei Haushalten mit IV oder AHV-Rente zu einer schlechteren Klassifikationsgüte sowie zu einer wesentlichen Überschätzung des Anteils einkommensschwacher Personen für diese Gruppe.

So fällt das auch bei der Datenbasis WiSiER mit Steuer gegenüber anderen Haushalten erhöhte Risiko von Einkommensschwäche bei Haushalten mit IV-Rentenbezug (25.8%) mit der Datengrundlage ohne Steuer (optimiert) nochmals wesentlich höher aus (31.9%), wie **Abbildung 10** zeigt. Das v.a. die BVG-beziehenden Haushalte fehlklassifiziert werden, zeigt sich im Streudiagramm in **Abbildung 15** im Anhang. Bei allfälligen Analysevorhaben wäre damit zusammenhängend zu beachten, dass in der Datenbasis Einkünfte aus Ergänzungsleistungen, anders als BVG-Renten, vorhanden sind. Entsprechende Zusammenhänge könnten dadurch fälschlicherweise gerade umgekehrt ausfallen: EL-Beziehende wären basierend auf dieser unvollständigen Datenbasis einkommensmässig meist besser gestellt als BVG-Beziehende, was in der Realität natürlich nicht so ist.

Der Anteil einkommensschwacher Personen in Haushalten mit EL wird mit den optimierten verknüpften Daten ohne Steuer gegenüber WiSiER mit Steuer um 4.3 Prozentpunkte unterschätzt (hier ohne Abbildung, vgl. separates Auswertungsdokument). Der Grund dafür liegt aber nicht in der unpräzisen Erfassung der Einkommen dieser Haushalte, sondern in der falschen Rangierung und Überschätzung der anderen Haushalte als Folge von deren lückenhaften Einkommensinformationen. Als rein relatives Mass ist P20 deutlich anfälliger für solche «Kontaminationseffekte» als die nur begrenzt relative Armutsgefährdungsgrenze.

Abbildung 10: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Haushalt mit IV-Rente**

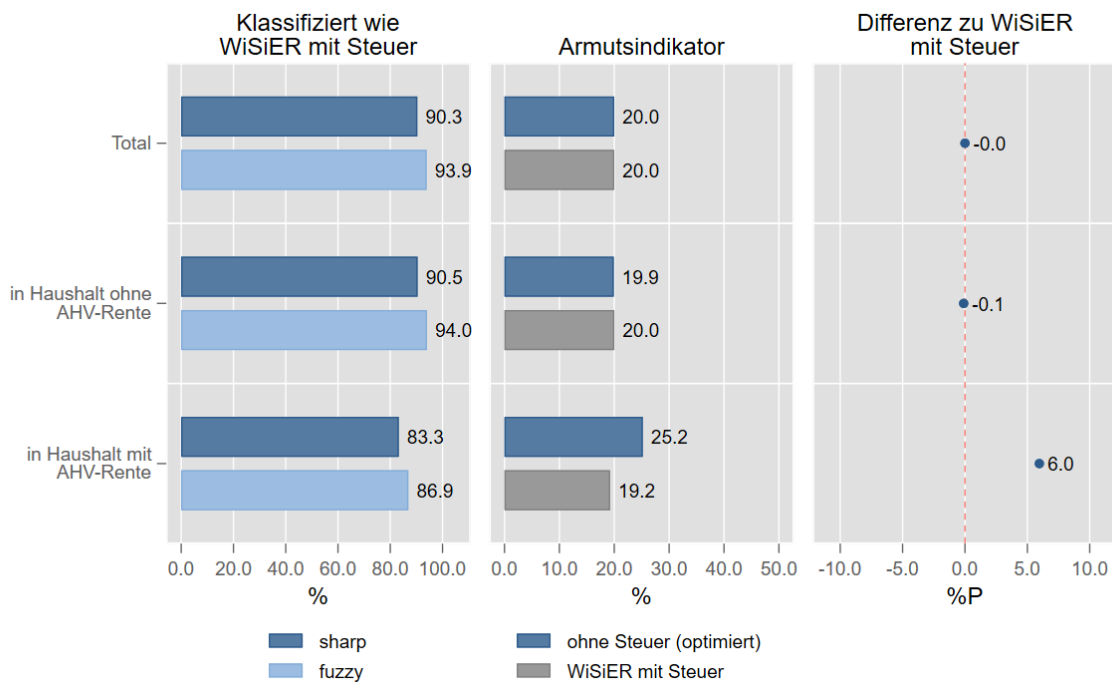


Quelle: WiSiER. Berechnungen BASS

4 Plausibilisierung der verknüpften Daten ohne Steuerinformationen zur Bestimmung der einkommensschwachen Bevölkerung

Aus dem gleichen Grund (fehlende Angaben zu BVG-Renteneinkünften) zeigt sich ein analoges Bild bei Haushalten mit AHV-Renten, wie in **Abbildung 11** abgebildet. Zu beachten ist, dass es sich dabei faktisch aufgrund der Altersobergrenze von 60 Jahren nicht um Alters-Rentner/innen, sondern um Haushalte mit Hinterlassenenrente handelt – Haushalte mit Witwen, Witwern oder Waisen. Sie weisen fälschlicherweise ein wesentlich erhöhtes Risiko der Einkommensschwäche gegenüber anderen Haushalten aus, obwohl dieses so generell zumindest nicht besteht, wie dies die WiSiER-Datenbasis zeigt (Abweichung um +6.0 Prozentpunkte).

Abbildung 11: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Haushalt mit AHV-Rente**

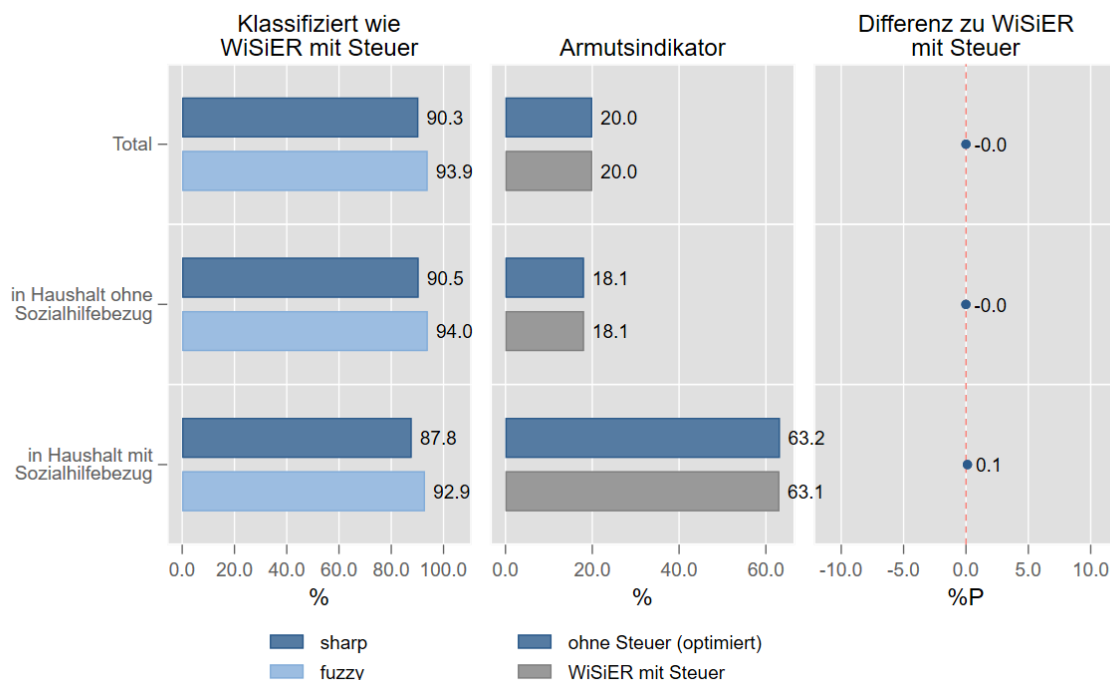


Quelle: WiSiER. Berechnungen BASS

Haushalte mit Sozialhilfebezug

Die gegenüber WiSiER mit Steuer konsistente Identifikation von einkommensschwachen Personen in Haushalten mit Sozialhilfebezug gelingt vergleichsweise zuverlässig und die Anteilswerte basierend auf den beiden Datenbasen sind quasi identisch. Was in der entsprechenden **Abbildung 12** auffällt, ist einzig der für beide Datengrundlagen relativ tiefe Wert von weniger als zwei Drittel der Sozialhilfebeziehenden, die einkommensschwach sind. Diesbezüglich könnte es aufschlussreich sein, in zukünftigen Analysen auch unterjährige Dynamiken zu beachten, da bekanntermassen ein wesentlicher Anteil der Sozialhilfebeziehenden diese Leistung weniger als ein Jahr lang bezieht, was dieses Ergebnis erklären könnte. Ein weiterer Faktor, welcher diesen auf den ersten Blick unplausiblen Befund erklärt, könnte auch der Umstand sein, dass mit WiSiER mit Steuer ja nicht das für einen Sozialhilfeanspruch relevante verfügbare Haushaltseinkommen erfasst wird und insbesondere die im Einzelfall unterschiedlich hohen Auslagen für Wohnkosten und Gesundheitsausgaben nicht berücksichtigt werden.

Abbildung 12: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Haushalt mit Sozialhilfebezug**



Quelle: WiSiER. Berechnungen BASS

Kurzbefunde zu weiteren Gruppierungsvariablen

Im separaten Auswertungsdokument finden sich Ergebnisse für weitere Gruppierungsvariablen, wie alternative Haushaltstypologien (u.a. nach Anzahl Kinder bei Paarhaushalten, analog zum BFS-Vorgehen), Geschlecht, Zivilstand, Staatsangehörigkeit oder Sprachregion. Diese sollen hier nur kurz angesprochen werden, ansonsten sei auf das separate Dokument verwiesen. Auch hier wird immer das Ergebnis für die optimierte verknüpfte Datenbasis ohne Steuerdaten herangezogen, sofern nicht explizit anders erwähnt.

■ **Zivilstand:** Die Einkommensschwäche/P20 von verwitweten und geschiedenen Personen wird wesentlich überschätzt (+8.4, respektive + 3.6 Prozentpunkte), im Einklang mit oben gezeigten Auswertungen und aus denselben Gründen – fehlende Angaben zu BVG-Renten sowie den Unterhaltszahlungen (siehe **Abbildung 13** unten).

■ **Staatsangehörigkeit/Aufenthaltsstatus:** Die Einkommensschwäche/P20 von Ausländerinnen und Ausländern, faktisch aufgrund der eingeschränkten Grundgesamtheit ohne Quellenbesteuerte (nur Personen mit Niederlassungsbewilligung C), wird unterschätzt (-2.2 Prozentpunkte Differenz).

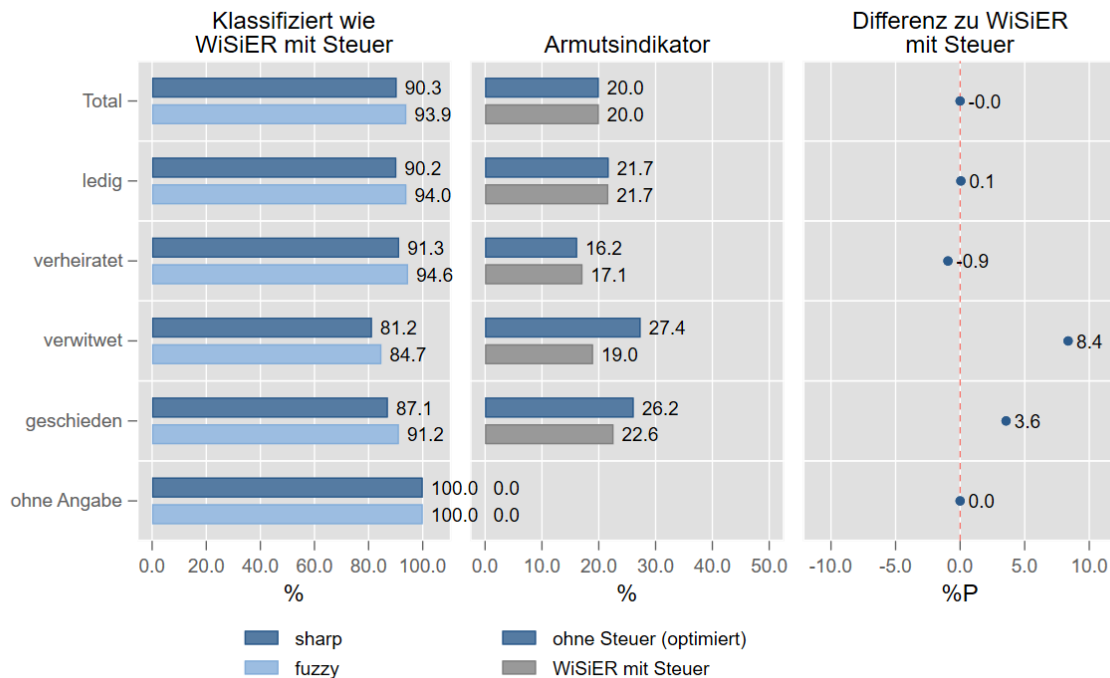
■ **Sprachregion:** Das allgemeine Muster/die Reihenfolge der Kategorien bleibt unverändert, allerdings wird das höhere Risiko für Einkommensschwäche bei Personen aus der lateinischen Schweiz gegenüber der Deutschschweiz noch zusätzlich überschätzt.

■ **Höchster Bildungsabschluss (Haushaltsebene):** Das allgemeine Muster/die Reihenfolge der Kategorien bezüglich Einkommensschwäche/P20 bleibt unverändert, die Differenzen sind relativ geringfügig. Auf Personenebene gilt dieser Befund ebenso.

■ **Haushalte mit Ergänzungsleistungen:** Das Muster eines höheren Risikos der Einkommensschwäche/P20 bei EL-Haushalten ist bei beiden Datengrundlagen identisch, allerdings wird der Anteil mit den verknüpften Daten ohne Steuer wesentlich unterschätzt (28.3 anstatt 32.6 mit WiSiER mit Steuer, Differenz von -4.3 Prozentpunkten). Der Grund dafür ist allerdings wie oben erwähnt indirekt: die Überschätzung der Einkommensschwäche bei einem Teil der nicht-EL-Haushalten, mehrheitlich aufgrund fehlender BVG-Renteneinkünfte.

Interessant ist, dass diese beobachtete Unterschätzung beim Indikator «60% Median» **nicht stattfindet**, da dieser **als nicht rein relatives Armutsmass robuster gegen solche «Kontaminationseffekte»** ist. Allerdings wird dort für beide Kategorien, EL-Haushalte wie nicht-EL-Haushalte, der Indikator um jeweils rund 2.5 Prozentpunkte überschätzt. Dieser Befund illustriert beispielhaft das unterschiedliche Verhalten der beiden Indikatoren aufgrund ihrer unterschiedlichen Berechnungsweise.

Abbildung 13: Plausibilisierung P20 – verknüpfte Daten ohne Steuer (optimiert), nach **Zivilstand**



Quelle: WiSiER. Berechnungen BASS

4.3 Schlussfolgerungen aus den Ergebnissen der Plausibilisierung

Die wichtigsten Erkenntnisse aus den Ergebnissen der Plausibilisierung lassen sich wie folgt zusammenfassen:

■ Bei der **verknüpften Datenbasis (ohne Steuer)** zeigt sich gegenüber der Referenzdatenbasis WiSiER mit Steuer ein **Anteil an abweichenden Klassifikationen bezüglich Einkommensschwäche (P20) im Ausmass von rund 10%**. Konkret bedeutet dies, dass eine von zehn Personen als (nicht) einkommensschwach klassifiziert wird, obwohl sie dies gemäss Referenzdatenbasis WiSiER (nicht) ist. Bei einer «fuzzy»-Klassifikation, welche knapp falsch zugeordnete Beobachtungen noch als korrekt behandelt, sind noch gut 6% der Individuen abweichend klassifiziert.

Problematisch ist bei diesen Fehlklassifikationen nicht ihr Ausmass – welches sich je nach Analysevorhaben in einem durchaus tolerierbaren Rahmen befindet –, als vielmehr der Umstand, dass diese Fehlklassifikationen **nicht zufällig, sondern systematisch** erfolgen (vgl. Folgeabschnitt).

■ Die systematischen Fehlklassifikationen manifestieren sich als **Verzerrungen, respektive Abweichungen bei den errechneten Anteilswerten** der einkommensschwachen Personen. Die gestützt auf die bekannten fehlenden Einkommenskomponenten im Voraus erwarteten stark von den systematischen Fehlklassifikationen betroffenen Bevölkerungssegmente bestätigten sich in der Plausibilisierung: **Einelternhaushalte, Haushalten mit Hinterlassenen sowie Haushalte mit IV-Rentenbezug** weisen eine im Vergleich zu anderen Gruppen wesentlich tiefere Klassifikationsgüte und in der Folge grössere Verzerrungen, respektive Abweichungen bei den Anteilswerten auf. Konkret kommt es in diesen Fällen zu einer Überschätzung des Anteils einkommensschwacher Personen. **Hauptursache sind die aufgrund der nicht verfügbaren Steuerdaten fehlenden Informationen** zu den für diese Bevölkerungssegmente besonders relevanten Einkommenskomponenten **BVG-Renteneinkünfte** und **Unterhaltszahlungen**.

Mit der Plausibilisierung konnten diese absehbaren Verzerrungen auch quantifiziert und somit ihr genaues Ausmass beziffert werden. Die Abweichungen bei den Anteilswerten sind für einzelne Gruppen wesentlich, wie oben bei den Ergebnissen der Plausibilisierung aufgezeigt wurde. Eine Verwendung dieses Indikators zur Einkommensschwäche ohne Weiteres basierend auf den verknüpften Daten ohne Steuer für Analysen im Rahmen des Armutsmonitorings scheint vor diesem Hintergrund fraglich. Die vorliegenden umfassenden Plausibilisierungsergebnissen ermöglichen zumindest, bei zukünftigen Analysen potentielle Verzerrungen und Limitationen genauer zu antizipieren, zu deklarieren und so eine bessere Einordnung der Ergebnisse vorzunehmen.

■ Ein **Ausschluss der aus methodischer Sicht problematischen Bevölkerungssegmente**, wie dies als mögliche Lösung gesehen werden könnte, ist **nicht zielführend**: Zum einen sind Alleinerziehende und IV-Rentner/innen bekannte Risikogruppen für Armut. Man würde also gerade jene Individuen aus den Analysen ausschliessen, die für die Armutsdebatte von besonderem Interesse sind. Zum anderen hätte ein Ausschluss bestimmter Gruppen schwer kontrollierbare Folgen auch auf weitere Analysen und deren Validität, da andere Kategorisierungen und deren Kategorien je nach ihrer Zusammensetzung mit Blick auf diese ausgeschlossenen Gruppen unterschiedlich betroffen, respektive verzerrt würden. Ein Beispiel: Die festgestellten Unterschiede zwischen Kantonen (über alle Bevölkerungsgruppen) könnten sich als Folge eines Ausschlusses von Personen in Haushalten von Alleinerziehenden wesentlich verändern, falls bestimmte Kantone einen wesentlich höheren Anteil an dieser Bevölkerungsgruppe aufweisen.

Hinweise auf zusätzliche relevante Bevölkerungssegmente, bei welchen es ebenfalls zu stärkeren Abweichungen bei den berechneten Anteilswerten kommt, abgesehen von den oben genannten, wurden im Rahmen der Plausibilisierung keine identifiziert. Zu erwähnen sind jedoch die bereits in

WiSiER bezüglich ihres Einkommens nur bedingt angemessen erfassten Bevölkerungssegmente. Diese umfassen namentlich **Haushalte von jungen Erwachsenen unter 25 Jahren**, bei welchen allfällige private Transferzahlungen des Elternhaushalts und allfällige Stipendienleistungen nicht erfasst werden, sowie **AHV-Selbständige**, bei welchen die Angaben zum Einkommen nur beschränkt valide sind. Nicht plausibilisiert werden konnte die Gruppe der **Ausländerinnen und Ausländer ohne Niederlassungsbewilligung C**, da diese mehrheitlich quellenbesteuert werden und somit nicht in den WiSiER-Daten mit Steuerinformationen abgedeckt sind. Da die in der verknüpften Datenbasis (ohne Steuer) zur Bestimmung des Erwerbseinkommens verwendeten IK-Daten auch für diese Personen verfügbar sind, kann diese Bevölkerungsgruppe mit der neuen Vorgehensweise nun ohne Einschränkungen in den Analysen mitberücksichtigt werden. Die **Gruppe der Ausländerinnen und Ausländer kann somit mit der verknüpften Datenbasis umfassender einbezogen werden**, als dies im Rahmen der WiSiER-Analysen gemacht wurde.

■ Bei den weiteren plausibilisierten Gruppierungsvariablen – abgesehen von den oben erwähnten Bevölkerungssegmenten mit grossen Verzerrungen – zeigten sich weniger ausgeprägte Differenzen bei den Anteilswerten. Die grundsätzlichen Muster und Reihenfolgen bezüglich des Risikos von Einkommensschwäche zeigen sich auch mit der Datenbasis ohne Steuerinformationen. Auch hier gibt es jedoch immer wieder vereinzelt wesentliche Differenzen gegenüber der Referenzdatenbasis WiSiER mit Steuer zu beobachten.

Wie sich die beobachteten wesentlichen Verzerrungen für bestimmte Gruppen über Kompositionseffekte auch in anderen, hier nicht untersuchten Auswertungskategorisierungen aktuell und zukünftig niederschlagen, ist schwierig zu antizipieren. Insbesondere Entwicklungen von Anteilswerten im zeitlichen Verlauf sind aufgrund von Kompositionseffekten anfällig auf solche Probleme.

Auch ist zu beachten, dass bei Kennzahlen zur Armut wie der Armutsquote auch geringfügige Veränderungen über die Zeit oder zwischen Bevölkerungsgruppen von auch nur einem oder wenigen Prozentpunkten substantiell relevant sind. Aus diesem Grund sind auch **bereits geringfügige Abweichungen problematisch**. Schwierig abzuschätzen ist ebenfalls, in welchem Ausmass sich die beobachteten Verzerrungen bei der **Analyse von bivariaten oder multivariaten Zusammenhängen** auswirken.

Dies heisst nicht, dass Auswertungen grundsätzlich nicht sinnvoll sind. Hingegen sind die entsprechenden Herausforderungen jeweils bei der Analyse zu reflektieren. Auch sind die Limitationen bei der Vermittlung der Befunde transparent auszuweisen.

■ Insgesamt zeigt die Plausibilisierung deutlich: **Die fehlende Verfügbarkeit von kantonalen Steuerinformationen verunmöglicht die Verwendung der verknüpften Administrativ- und Befragungsdaten zur Analyse von Armut in der Schweiz nicht, schmälert deren Nutzen für das Armutsmonitoring jedoch wesentlich**. Ein Auftrag an das nationale Armutsmonitoring ist die Durchführung kantonal vergleichender Analysen. Um Kennzahlen wie kantonale Armutsquoten vergleichend darzustellen, sind Steuerdaten notwendig. Die **Verfügbarkeit von kantonalen Steuerdaten** bleibt vor diesem Hintergrund ein **zentrales und dringendes Erfordernis für valide Analysen zur Armut auf kantonaler Ebene**. Falls die Steuerdaten zu einem späteren Zeitpunkt zur Verfügung stehen, so ist darauf zu achten, dass dies möglichst auch für rückwärtige Jahre gilt, um die Analysen von längeren Zeitreihen zu einem möglichst frühen Zeitpunkt zu ermöglichen.

■ Die vorgenommenen **Optimierungen** der verknüpften Daten ohne Steuer, konkret die Modellierung des Nettolohns und der Kinderzulagen, hat einen **qualitätssteigernden Effekt auf die Ergebnisse, dieser fällt allerdings sehr gering aus**, teilweise ist er nicht sichtbar. Da der Aufwand hierfür

nicht allzu gross ist, ist aus unserer Sicht deshalb diese Optimierung beizubehalten. Das beschränkte Optimierungspotenzial verdeutlicht zugleich, dass die Folgen der grossen Lücken bei den Einkommenskomponenten nicht einfach mit der Modellierung oder der zusätzlichen Integration in den verknüpften Datensatz von Informationen zu sekundären Einkommenskomponenten, wie etwa den Kinderzulagen, kompensiert werden können. Relevant sind letztlich primär jene Einkommenskomponenten, die substantiell zum Haushaltseinkommen beitragen, wie etwa Alimente oder BVG-Rentenzahlungen.

■ Die **alternative Masszahl «60% des Medians»** als Indikator unterscheidet sich von P20 aufgrund ihrer Berechnungsweise, was wiederum zu unterschiedlichen Ergebnissen bei der Plausibilisierung führt. P20 ist als rein relative Kennzahl anfälliger auf «Kontaminationseffekte» von stark und systematisch fehlklassifizierten Kategorien auf andere Kategorien, wie oben aufgezeigt wurde. Die Masszahl «60% des Medians» weist wiederum aufgrund der Berechnungsweise bereits im Total Differenzen zwischen der Referenzdatenbasis WiSiER mit Steuer gegenüber der Datengrundlage ohne Steuer auf, die bei P20 per Definition nicht entstehen können. Dafür sind bei der Masszahl «60% Median» die Ergebnisse auch beim Total von substantiellem Interesse, weil sie variieren können (bspw. ist die Beobachtung im zeitlichen Verlauf auch im Total möglich). Diese spezifischen Vor- und Nachteile gilt es abzuwägen bezüglich des Entscheids, welcher Indikator für bestimmte Fragestellungen zur Anwendung kommen soll. Aus unserer Sicht gibt es hier keinen klaren Gewinner, letztlich hängt dies davon ab, wie man für eine spezifische Fragestellung die jeweiligen Vor- und Nachteile gewichtet.

■ Die Variante des Indikators **Datenbasis «nur Erwerbseinkommen»** (inkl. allfälliger ALV-Leistungen), welcher bewusst auf die primäre Einkommensquelle Erwerb fokussiert, zeigte bei der durchgeführten Plausibilisierung wie erwartet stärkere Abweichungen von der Referenzdatenbasis WiSiER mit Steuer als die verknüpften Daten ohne Steuer aufgrund der bei den letzteren zusätzlich berücksichtigten Einkommenskomponenten (bedarfsabhängige Sozialleistungen und Sozialversicherungsleistungen). Die relevante Frage mit Blick auf die Verwendung dieses alternativen Indikators ist, wie der Mehrwert einer klaren Konzeptualisierung und präzisen Messung des Konstrukts «Erwerbseinkommen» gegenüber der Fähigkeit, einkommensschwache Personen korrekt zu identifizieren, gewichtet wird. Bei dieser Datenbasis «nur Erwerbseinkommen» sind die Abweichungen gegenüber WiSiER mit Steuerdaten nochmals etwas grösser als bei den verknüpften Daten ohne Steuer, das grundsätzliche Muster bezüglich der Risikogruppen bleibt aber auch hier erhalten.

Der Entscheid, ob diese P20-Indikatoren-Variante «Erwerbseinkommen» in ausgewählten Analysen zur Anwendung kommen soll, ist letztlich nicht allein aufgrund des Abschneidens in der Plausibilisierung, als vielmehr auch nach konzeptuellen Überlegungen zu treffen – u.a. mit Blick auf die Eignung für die zu beantwortende Fragestellungen, der Validität der Ergebnisse in Bezug auf das zu messende Konstrukt und der Belastbarkeit der Ergebnisse. Steht beispielsweise nicht Armut generell, sondern aufgrund der vorrangigen Bedeutung des Erwerbseinkommens als primäre Einkommensquelle von Personen im Erwerbsalter spezifisch die **Erwerbseinkommensschwäche** im Fokus einer Fragestellung, so sind mit diesem Indikator sowohl valide als auch aufschlussreiche Analysen durchführbar.

5 Literaturverzeichnis

- Bio Tchané, Abdoulaye (2019): Understanding the P20: A new approach to characterising and addressing poverty. OECD. Paris. Online: <https://www.oecd.org/dac/understandingthep20anewapproachtocharacterisingandaddressingpoverty.htm>.
- Bischof, Severin, Tabea Kaderli, Lena Liechti, Jürg Guggisberg (2023). Die wirtschaftliche Situation von Familien in der Schweiz - Bedeutung von Geburten sowie Trennungen und Scheidungen. Hg. Bundesamt für Sozialversicherungen (Beiträge zur sozialen Sicherheit, 01/23). Büro BASS, Bern.
- Braun-Dubler, Nils, Vera Frei, Tabea Kaderli, Florian Roth (2022): Wer geht wann in Rente? Ausgestaltung und Determinanten des Rentenübergangs. Hg. Bundesamt für Sozialversicherungen (Forschungsbericht 05/22.1).
- Brülle, J., & Gangl, M. (2023). Verfestigung von Armut und die zunehmende Bedeutung von Pfadabhängigkeiten im Lebenslauf. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 1-35.
- Bundesamt für Sozialversicherungen BSV (2021). Konzept für ein Nationales Armutsmonitoring in Erfüllung der Motion 19.3953 WBK-S.
- Bundesamt für Sozialversicherungen BSV (2023a). Kurzfassung zum Detailkonzept des nationalen Armutsmonitorings.
- Bundesamt für Sozialversicherungen BSV (2023b). Statistische Definition Selbständigerwerbende. Bern.
- Bundesamt für Statistik BFS (2021). Berechnung eines neuen Erwerbseinkommens anhand der ZAS- und SE-Daten. Deskriptive Analyse der Unterschiede zum LSE-Lohn. Reihe «Experimental Statistics». Neuchâtel.
- Drilling, Matthias (2004). Soziale Ungleichheit und Armut in der Schweiz: Zum Stand der Forschung. Download zu Drilling, Matthias (2004), Young urban poor. Abstiegsprozesse in den Zentren der Sozialstaaten. Wiesbaden: VS-Verlag.
- Fluder, Robert, Hümbelin, Oliver, Luchsinger, Larissa, Richard, Tina (2020): Ein Armutsmonitoring für die Schweiz: Modellvorhaben am Beispiel des Kantons Bern. Caritas und Berner Fachhochschule BFH.
- Guggisberg, Jürg, Severin Bischof, Roman Liesch, Melania Rudin/Büro BASS (2020). Sozialhilfebezug in der Mehrjahresperspektive und im Lebensverlauf. Hg. Bundesamt für Sozialversicherungen (Forschungsbericht 05/20).
- Guggisberg, Jürg, Hugo Bodory, Dominic Höglinger, Severin Bischof, Melania Rudin (2020). Gesundheit der Migrationsbevölkerung – Ergebnisse der Schweizerischen Gesundheitsbefragung 2017. Im Auftrag des Bundesamts für Gesundheit (BAG). Büro BASS, Bern.
- Hümbelin, O., Farys, R., Jann, B. & Lehmann, O. (2021). Umverteilung über Steuern und Sozialleistungen in der Schweiz-. Social Change in Switzerland, N°28.
- Liechti, David, Michael Morlok, Michael Siegenthaler (2020). Situation, Entwicklung und Auswirkungen der Langzeitarbeitslosigkeit. Studie im Auftrag der Aufsichtskommission für den Ausgleichsfonder der Arbeitslosenversicherung. SECO Publikation Arbeitsmarktpolitik Nr. 59 (02.2020). B,S,S Volkswirtschaftliche Beratung & KOF Konjunkturforschungsstelle ETH Zürich.
- United Nations Economic Commission for Europe (2011). Canberra group handbook on household income statistics.

- Wanner, Philippe (2019). Préparation d'une base de données sur la situation économique des personnes en âge d'activité et à l'âge de la retraite (WiSiER). Hg. Bundesamt für Sozialversicherungen (Beiträge zur sozialen Sicherheit, 04/19).
- Wanner, Phillipe, Roxanne Gerber (2022). Die wirtschaftliche Situation der Bevölkerung im Erwerbs- und im Rentenalter. Hg. Bundesamt für Sozialversicherungen (Beiträge zur sozialen Sicherheit, 04/22).

Anhang

A-1 Ergänzende Tabellen und Abbildungen

Konfusionsmatrix und abgeleitete Kenngrössen

Tabelle 3: Konfusions-Matrix und Kenngrössen der Klassifikation P20 mittels Datenbasis Verknüpfte Daten ohne Steuer gegenüber Referenzdatenbasis WiSiER mit Steuer

True	Classified		Total
	+	-	
+	371354	118031	489385
-	118023	1839469	1957492
Total	489377	1957500	2446877

Measure		Estimate
Sensitivity	$A/(A+B)$	75.88%
Specificity	$D/(C+D)$	93.97%
Positive predictive value	$A/(A+C)$	75.88%
Negative predictive value	$D/(B+D)$	93.97%
False positive rate	$C/(C+D)$	6.03%
False negative rate	$A/(A+B)$	24.12%
Correctly classified	$A+D/(A+B+C+D)$	90.35%
ROC area		84.93%

Quelle: WiSiER. Berechnungen BASS

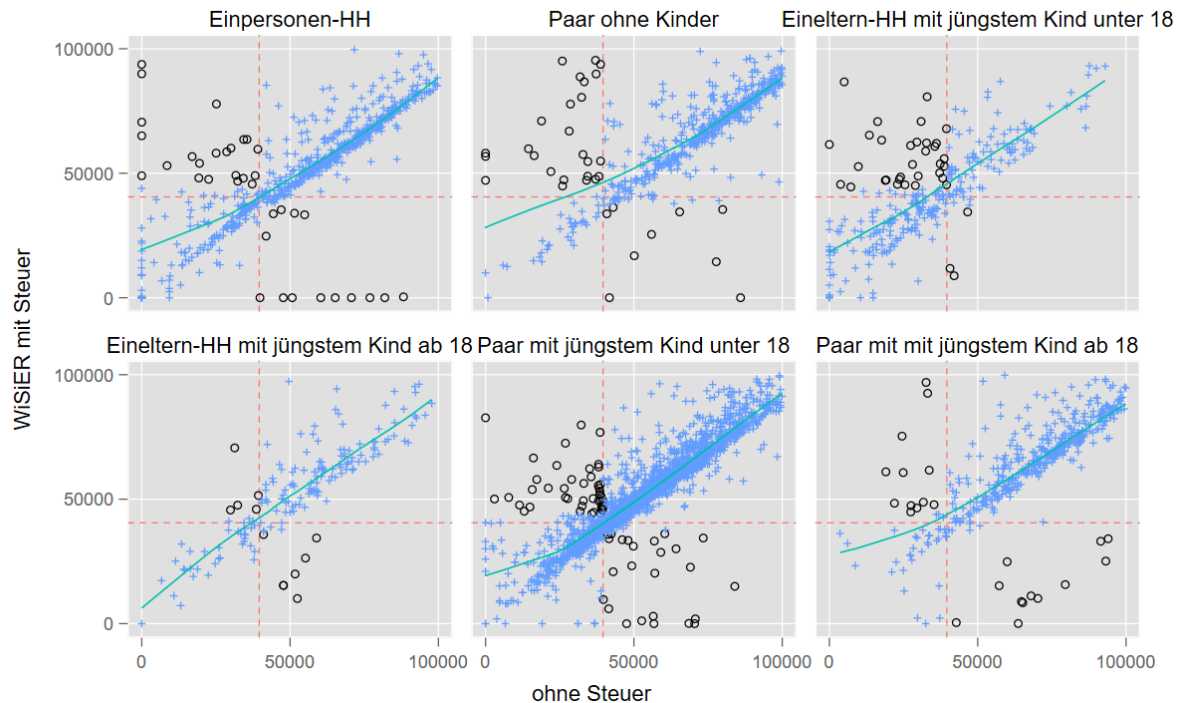
In **Tabelle 3** sind die Konfusions-Matrix für die P20-Klassifikation durch die verknüpften Daten mit Steuer gegenüber der Referenzdatenbasis WiSiER mit Steuer sowie die daraus abgeleiteten Kenngrössen ausgeführt. In den Abbildungen ist jeweils nur die zentrale Kenngrösse der korrekt klassifizierten Beobachtungen abgebildet. In der Tabelle sind weitere Kenngrössen von sekundärem Interesse ersichtlich, namentlich:

■ **Sensitivität:** Der Anteil korrekt als einkommensschwach klassifizierter Personen an allen gemäss Referenzdatenbasis WiSiER mit Steuer tatsächlich einkommensschwachen Personen, d.h. wie gut positive Beobachtungen erkannt werden. Dieser Wert beträgt 75.9 Prozent.

■ **Spezifität:** Der Anteil korrekt als nicht einkommensschwach klassifizierter Personen an allen gemäss Referenzdatenbasis WiSiER mit Steuer tatsächlich nicht einkommensschwachen Personen, d.h. wie gut negative Beobachtungen erkannt werden. Dieser Wert beträgt 94.0 Prozent.

Streudiagramm zur Visualisierung des Vorgehens bei der «fuzzy»-Klassifikation

Abbildung 14: Äquivalenz-Haushaltseinkommen WiSiER mit Steuer vs. verknüpfte Daten ohne Steuer mit Klassifikation «fuzzy» nach Haushaltstyp (mit Alter jüngstes Kind).



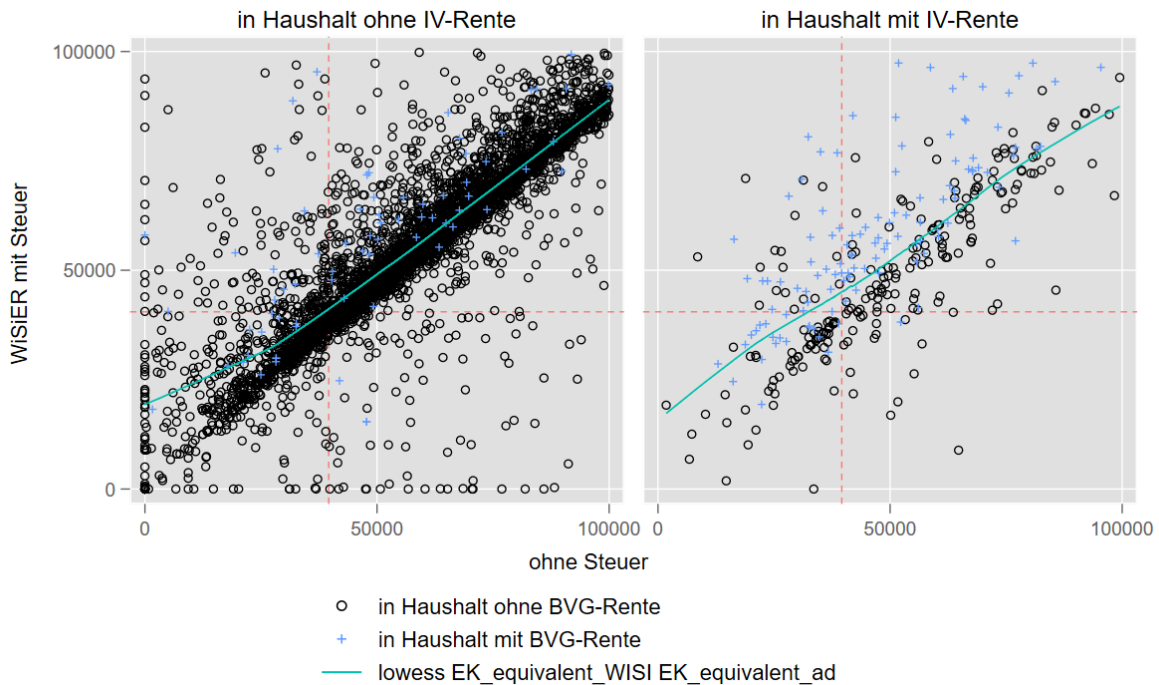
Rot gestrichelte Linie: P20-Perzentilsgrenze. Korrekt klassifizierte Personen («fuzzy»-Klassifikation) sind mit blauem Kreuz, fehlerhafte Klassifikationen mit schwarzem Kreis ausgewiesen. Ohne andere und nicht bestimmbare Haushaltstypen. Abbildung basiert auf einer Stichprobe zur besseren Darstellung. Quelle: WiSiER. Berechnungen BASS

In **Abbildung 14** wird ersichtlich, wie die Bestimmung der Güte der «fuzzy»-Klassifikation der Personen in «einkommensschwach» (P20) oder «nicht einkommensschwach» (über Grenzwert P20) erfolgt: Die jeweiligen P20-Grenzwerte für die Referenzdatenbasis WiSiER mit Steuer als auch die verknüpften Daten ohne Steuer sind als rote Linien eingetragen, mit ihnen ergeben sich die vier Quadranten der Klassifikationsmatrix. Blau markierte Personen werden als gleich («korrekt») klassifiziert eingestuft (Quadrant unten links und Quadrant oben rechts), wobei es bei der Referenzdatenbasis WiSiER mit Steuer eine gewisse Bandbreite gibt (15. Perzentil bis 25. Perzentil, +/- 5 Prozentpunkte), in welcher aufgrund der knappen Klassifikation mittels WiSiER mit Steuer eine Zuweisung durch die Prüfdatenbasis Verknüpfte Daten ohne Steuer zu beiden Gruppen als korrekt betrachtet wird.

Visuell deutlich wird in der Abbildung auch die beträchtliche Verzerrung bei den Eineltern-HH mit jüngstem Kind unter 18 Jahren, die sich in einer starken Häufung von Fehlklassifikationen (schwarze Kreise) einseitig im oberen linken Quadranten (und kaum im unteren rechten Quadranten) manifestiert.

Streudiagramme zu ausgewählten Ergebnissen

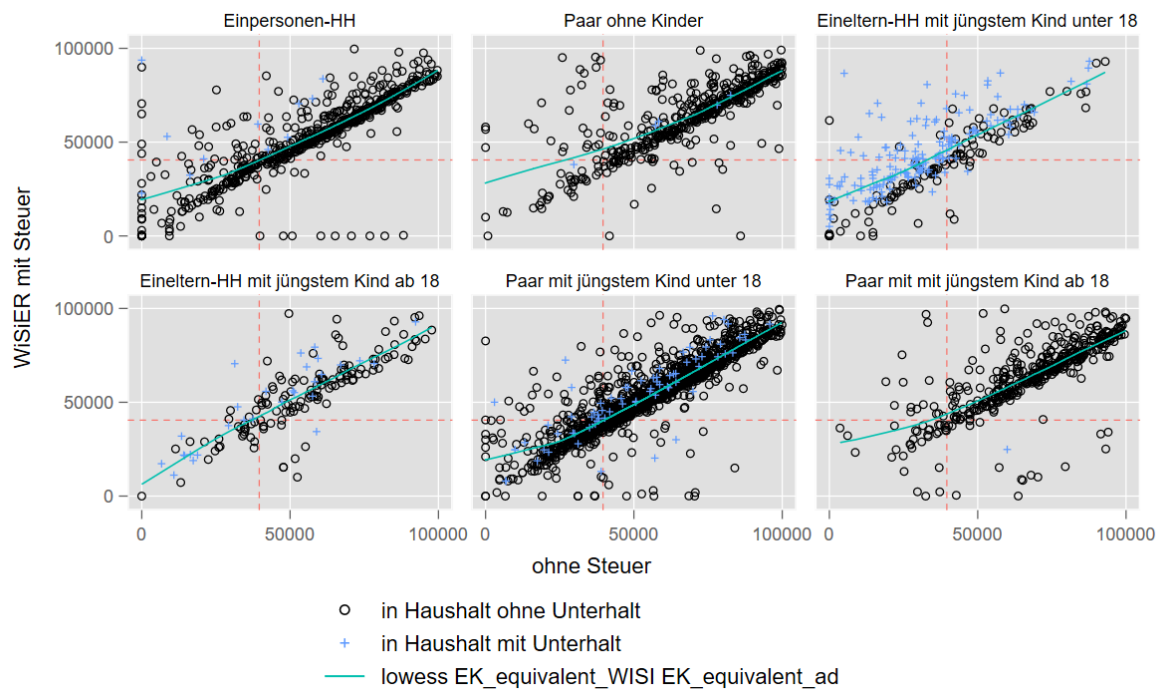
Abbildung 15: Äquivalenz-Haushaltseinkommen WiSiER mit Steuer vs. verknüpfte Daten ohne Steuer nach Haushalt ohne und mit IV Rente, markiert falls zusätzliche BVG-Rente.



Ohne andere und nicht bestimmbare Haushaltstypen. Abbildung basiert auf einer Stichprobe zur besseren Darstellung.
Quelle: WiSiER. Berechnungen BASS

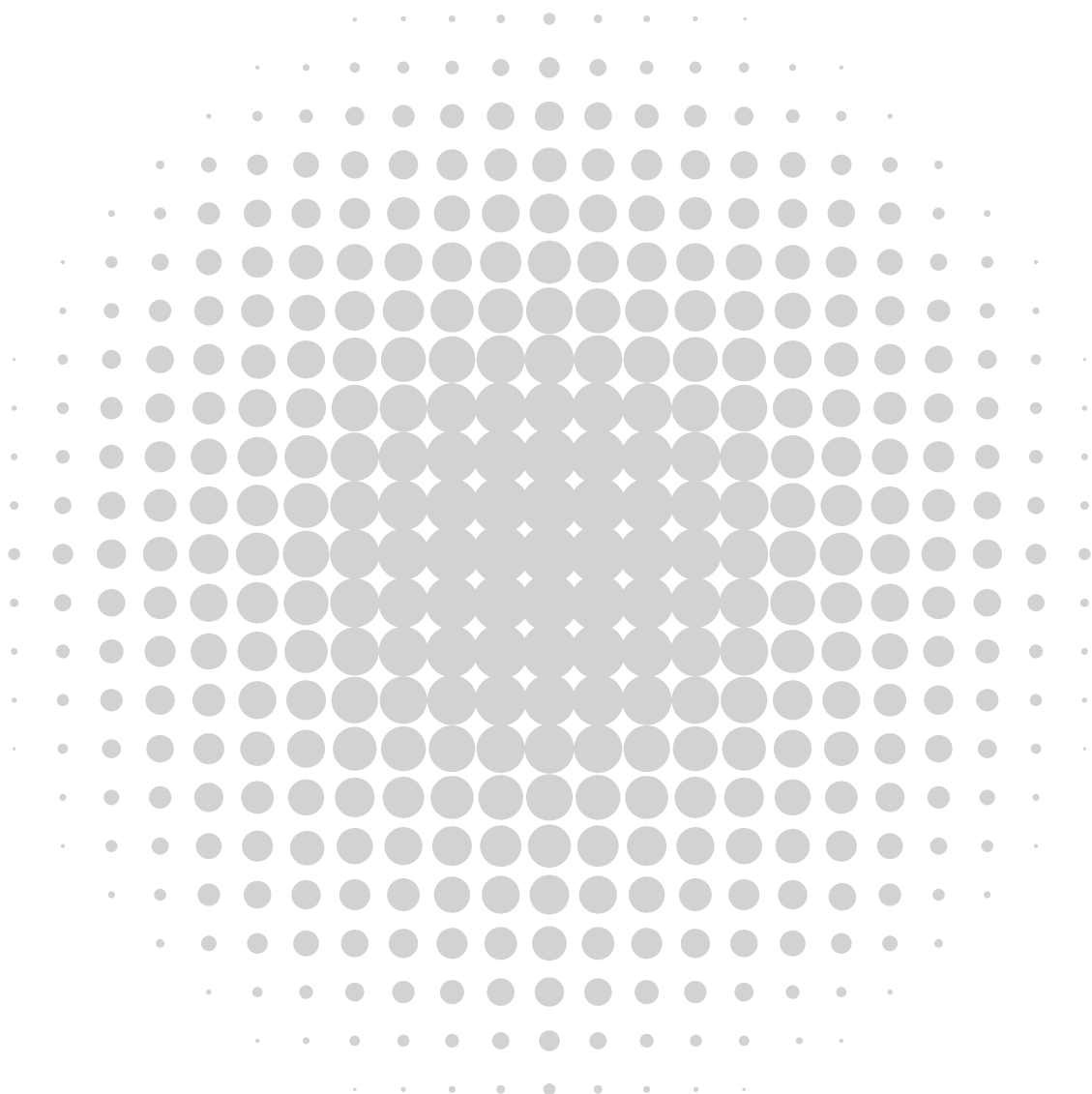
In **Abbildung 15** ist ersichtlich, dass Haushalte mit IV-Rente (rechte Grafik), welche zusätzlich eine BVG-Rente erhalten (blaues Kreuz), sich gehäuft als Fehlklassifizierung im linken oberen Quadranten befinden (und es wenig entsprechende «ausgleichende» Fehlklassifizierungen im unteren rechten Quadranten gibt, wie dies bei zufälligen Fehlklassifikationen zu erwarten wäre). Die LOESS-Kurve zeigt zudem, dass der Zusammenhang von der idealen Erwartung abweicht (Diagonale durch den Nullpunkt und den Schnittpunkt der beiden roten Linien mit den P20-Grenzwerten)

Abbildung 16: Äquivalenz-Haushaltseinkommen WiSiER mit Steuer vs. verknüpfte Daten ohne Steuer nach Haushaltstyp (mit Alter jüngstes Kind), markiert falls Unterhaltszahlung.



Ohne andere und nicht bestimmbare Haushaltstypen. Abbildung basiert auf einer Stichprobe zur besseren Darstellung.
Quelle: WiSiER. Berechnungen BASS

In **Abbildung 16**, Grafik oben rechts, ist ersichtlich, dass Eineltern-Haushalte mit jüngstem Kind unter 18, welche zusätzlich Unterhaltszahlungen erhalten (blaues Kreuz), sich gehäuft im linken oberen Quadranten befinden – und somit systematisch fehlklassifiziert sind. Die entsprechende LOESS-Kurve zeigt zudem, dass der Zusammenhang von der idealen Erwartung abweicht (Diagonale durch Nullpunkt und Schnittpunkt der beiden roten Linien mit den P20-Grenzwerten)



bsv.admin.ch



Forschungspublikationen – «Beiträge zur Sozialen Sicherheit»