

Dipl.-Kauffrau (FH) Julia Vitlina, Dipl.-Sozialwissenschaftlerin Urszula Sikorski,
Dipl.-Wirtschaftsingenieur Hartmut Minkel

Generierung der Einkommensvariablen in der deutschen EU-SILC-Erhebung „LEBEN IN EUROPA“

Die Europäische Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC¹), die in Deutschland unter dem Namen *LEBEN IN EUROPA* durchgeführt wird, stellt die zentrale Datenbasis für die Europäische Sozialberichterstattung dar. Die Einkommensdaten liefern dabei die wichtigste Grundlage zur Berechnung der vom Europäischen Rat verabschiedeten und für alle Mitgliedstaaten der Europäischen Union (EU) vergleichbaren Indikatoren zur Messung von Armut und sozialer Ausgrenzung.

Inzwischen nehmen 31 Länder an EU-SILC teil: die 27 Mitgliedstaaten der EU sowie Kroatien, Island, Norwegen und die Schweiz. Um eine EU-weite Vergleichbarkeit der EU-SILC-Ergebnisse zu gewährleisten und zugleich die sehr unterschiedlichen und komplexen nationalen institutionellen Gegebenheiten adäquat abzubilden, sind in den Rechtsgrundlagen von EU-SILC zahlreiche methodische, definitorische und zeitliche Vorgaben zur Erhebung und zur Generierung der EU-SILC-Daten festgelegt.

Einer der zeitlich und inhaltlich anspruchsvollsten Arbeitsschritte bei der Generierung von vergleichbaren EU-SILC-Daten ist die Aufbereitung der nationalen Einkommensinformationen. Dieser Aufbereitungsprozess wurde seit der ersten Befragungswelle 2005 in Deutschland kontinuierlich inhaltlich, methodisch und technisch weiterentwickelt. Der Aufsatz gibt einen Einblick in den aktuellen Prozess der Aufbereitung der Einkommensinformationen in der deutschen EU-SILC-Erhebung, und zwar von der Plausibilisierung der erfassten Daten über Imputationen bis zur Generierung und Prüfung der finalen Einkommenszielvariablen.

Das Einkommenskonzept von EU-SILC

Das Einkommenskonzept von EU-SILC basiert auf den Empfehlungen der von der Statistischen Kommission der Vereinten Nationen aufgerufenen internationalen Experten-Gruppe zur Haushaltsstatistik „Canberra Group“², die im Jahr 2001 Richtlinien für vergleichbare Haushaltsstatistiken festgelegt hat. Die Erfassung der Sozialleistungen folgt dem Europäischen System der Integrierten Sozialschutzstatistik (ESSOSS) von 1996³. Das Konzept definiert das Einkommen als „regelmäßige Einkünfte wie Löhne und Gehälter, Einkommen aus selbstständiger Arbeit, Zins- und Dividenden erträge aus angelegten Finanzmitteln, Renten oder andere Leistungen der Sozialversicherung und sonstige laufend empfangene Transferleistungen. Große und unregelmäßige Einkünfte aus Vermächtnissen und dergleichen werden als Vermögenstransfers betrachtet, da es unwahrscheinlich ist, dass sie sofort dem Konsum zugeführt werden, und es sich um einmalige Zahlungen handelt“⁴. Das Einkommenskonzept umfasst sowohl monetäre als auch nicht monetäre Leistungen. Zu den nicht monetären Einkommen gehören zum Beispiel Einkommen aus unselbstständiger Tätigkeit in Form von Sachleistungen oder die für den Eigenverbrauch produzierten Waren. Um das Einkommenskonzept auf der europäischen Ebene harmonisiert umsetzen zu können, wird für jede Einkommenszielvariable eine genaue Definition festgelegt. Je nach Komplexität der Zielvariablen müssen die Einkommensfragen unter Berücksichtigung der

1 European Union Statistics on Income and Living Conditions.

2 The Canberra Group: “Expert Group on Household Income Statistics. Final Report and Recommendations”, Ottawa 2001, Genf 2011 (Second Edition).

3 Siehe Europäische Kommission: „ESSOSS-Handbuch“, Luxemburg 1996; siehe auch “ESSPROS Manual – The European System of integrated Social Protection Statistics. 2011 Edition”, Luxemburg 2011.

4 Siehe Fußnote 2.

nationalen Gegebenheiten entsprechend detailliert in den nationalen Fragebogen umgesetzt werden. Um den aktuellen Lebensstandard eines Haushaltes zu erfassen, werden laufende Einkommen über den Berichtszeitraum eines ganzen Jahres erhoben.

Die wichtigsten Einkommensvariablen sind

- › das Haushaltsbruttoeinkommen und
- › das verfügbare Haushaltsnettoeinkommen (Haushaltsbruttoeinkommen nach Abzug geleisteter laufender Transfers, unter anderem von Einkommensteuer und Sozialversicherungsbeiträgen).

Bei der Generierung der Zielvariablen nach den Vorgaben von Eurostat werden unterschiedliche Einkommensbestandteile unter anderem zum Haushaltsbrutto- beziehungsweise Haushaltsnettoeinkommen aggregiert. In der Übersicht 1 sind diese Einkommensaggregate mit ihren Bestandteilen in Form von Zielvariablen dargestellt. Aus der Bezeichnung jeder generierten Einkommenszielvariablen ist zu erkennen, ob es sich um ein Haushalts- oder ein Personeneinkommen (HY-/PY-Variable) handelt. Zu jeder EU-SILC-Einkommenszielvariablen muss eine sogenannte Flag-Variable gebildet werden, in der unter anderem Informationen über Aufbereitungsmethode und Imputationsumfang abgelegt sind.

Übersicht 1

Zusammensetzung des Haushaltsbrutto- und Haushaltsnettoeinkommens in EU-SILC 2010

EU-SILC-Einkommenszielvariable	Bezeichnung
Summe der Einkommensunterkomponenten auf der Personenebene	
Bruttoeinkommen aus unselbstständiger Tätigkeit in Form von Geld- oder geldwerten Leistungen	PY010G
+ Firmenwagen	PY021G
+ Bruttogewinn oder -verlust aus selbstständiger Tätigkeit	PY050G
+ Arbeitslosenunterstützung	PY090G
+ Altersleistungen	PY100G
+ Hinterbliebenenleistungen	PY110G
+ Krankengeld	PY120G
+ Invaliditätsleistungen	PY130G
+ Ausbildungsbezogene Leistungen	PY140G
Einkommensunterkomponenten auf der Haushaltsebene	
+ Einkommen aus Vermietung oder Verpachtung	HY040G
+ Familienleistungen/Kindergeld	HY050G
+ Sonstige Leistungen gegen soziale Ausgrenzung	HY060G
+ Wohnungsbeihilfen	HY070G
+ Regelmäßig erhaltene Geldtransfers zwischen privaten Haushalten	HY080G
+ Zinsen, Dividenden, Gewinne aus Kapitalanlagen in Unternehmen ohne eigene Rechtspersönlichkeit	HY090G
+ Von Personen unter 16 Jahren bezogenes Einkommen	HY110G
= Haushaltsbruttoeinkommen	HY010
- Regelmäßige Vermögensteuern	HY120G
- Regelmäßig geleistete Geldtransfers zwischen privaten Haushalten	HY130G
- Einkommensteuern und Sozialbeiträge	HY140G
= Verfügbares Haushaltsnettoeinkommen	HY020

Die Messung des Einkommens in der deutschen EU-SILC-Erhebung

EU-SILC wird in Deutschland seit 2005 jährlich als schriftliche Befragung LEBEN IN EUROPA durchgeführt.⁵ In Deutschland nehmen jährlich etwa 14 000 Haushalte an der Befragung teil, von denen jedes Jahr ein Viertel ausgetauscht wird. Die Haushalte erhalten einen Haushaltsfragebogen und jedes Haushaltsmitglied über 16 Jahren einen Personenfragebogen. Die Erhebungsunterlagen enthalten neben Fragen zum Bildungsabschluss, zur Erwerbssituation, zum Gesundheitszustand oder zur Wohnsituation detaillierte Fragen zum Einkommen. Der Referenzzeitraum für die Einkommensfragen ist in Deutschland das der Erhebung jeweils vorangegangene Kalenderjahr (die Einkommensangaben aus der EU-SILC-Erhebung 2010 beziehen sich also auf das Jahr 2009). Auf diese Weise werden sämtliche im Laufe eines Jahres bezogenen verschiedenen Einkommensarten berücksichtigt, um die Einkommenssituation eines Haushalts zu ermitteln.⁶

Der Personenfragebogen enthielt zum Beispiel im Jahr 2010 etwa 60 Fragen zu personenbezogenen Einkommenskomponenten wie

- › Bruttoeinkommen aus unselbstständiger Tätigkeit (Höhe des Lohns/des Gehalts aus der Haupt-/Nebenerwerbstätigkeit, Sonderzahlungen, Sach- und Naturalleistungen),
- › Bruttoeinkommen aus selbstständiger Tätigkeit,
- › Höhe der Renten/Pensionen,
- › Höhe von staatlichen Leistungen (Leistungen von der Bundesagentur für Arbeit/von Jobcentern beziehungsweise von anderen öffentlichen Trägern, wie unter anderem Arbeitslosengeld I, Pflegegeld oder Elterngeld, Leistungen aus privater Pflegezusatz- oder Krankentagegeldversicherung),
- › Höhe der gezahlten Steuern und Sozialversicherungsbeiträge sowie
- › Höhe der Beiträge für private Vorsorgeformen.

Der Haushaltsfragebogen enthält Fragen zur Höhe jener Einkommenskomponenten, für die unterstellt wird, dass sie vom gesamten Haushalt bezogen werden beziehungsweise die in der Regel nicht eindeutig einem bestimmten Haushaltsmitglied zugeordnet werden können, wie

⁵ Rechtsgrundlage für die Erhebung ist die Verordnung (EG) Nr. 1177/2003 des Europäischen Parlaments und des Rates vom 16. Juni 2003 für die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC) (Amtsblatt der EU Nr. L 165, Seite 1), geändert durch die Verordnung (EG) Nr. 1553/2005 des Europäischen Parlaments und des Rates vom 7. September 2005 (Amtsblatt der EU Nr. L 255, Seite 6).

⁶ Wichtig ist, dass der Haushalt nicht nur nach der Höhe des monatlichen Haushaltsnettoeinkommens gefragt wird (sogenannte Income-Screener-Frage), denn der Haushalt „vergisst“ hier häufig jene Einkommenskomponenten mit zu berücksichtigenden, die zum Beispiel nicht monatlich bezogen werden, deren Höhe aber einen signifikanten Einfluss auf das Wohlfahrtsniveau des Haushalts haben kann (zum Beispiel Prämienzahlungen, Kapitaleinnahmen und so weiter). Bei der Income-Screener-Frage wird also vor allem die Einkommenssituation jener Haushalte unterschätzt, die über mehrere verschiedene Einkommensquellen verfügen und/oder diese Einkommensarten nicht jeden Monat beziehen.

- › Einkommen aus öffentlichen Zahlungen (Kindergeld, Kinderzuschlag),
- › Einkommen aus privaten Zahlungen von nicht im Haushalt lebenden Personen,
- › Einkommen aus Vermietung und Verpachtung,
- › Einnahmen aus Wertanlagen,
- › Einkommen aus eigenem Garten oder eigener Kleintierhaltung,
- › Höhe der Grundsteuer für Grundbesitz,
- › Höhe der geleisteten Zahlungen an Personen, die nicht im Haushalt leben,
- › Höhe der Zahlungen für Kredite für eine selbst bewohnte Wohnung oder ein selbst bewohntes Haus.

Darüber hinaus enthält der Haushaltsfragebogen Fragen zu Einkommensarten, die alle Haushaltsmitglieder betreffen, wie

- › staatliche Leistungen zum Lebensunterhalt (unter anderem Arbeitslosengeld II, Grundsicherung, Sozialgeld, Sozialhilfe, Wohngeld in Form von Miet- oder Lastenzuschuss, Kosten der Unterkunft und so weiter),
- › Höhe von Einkommensteuerrückerstattungen beziehungsweise -nachzahlungen.

Einkünfte von Kindern unter 16 Jahren werden ebenfalls im Haushaltsfragebogen abgefragt.

Zusätzlich zur Höhe der Einkünfte wird der jeweilige Bezugszeitraum erfasst. Der Bezugszeitraum wird je nach Einkommensart unterschiedlich erfragt, zum Beispiel werden Lohn beziehungsweise Gehalt aus unselbstständiger Arbeit als Jahresbetrag erfragt, der Bezug von Arbeitslosengeld II als Monatsbetrag. Zusätzlich wird bei Bezug von Arbeitslosengeld II auch die Anzahl der Monate erfragt, in denen diese staatliche Leistung bezogen wurde. Damit die Haushalte möglichst alle ihre Einkommen vollständig mitteilen, können die Haushalte oder Personen in zusätzlichen Klartextfeldern Einkommensarten angeben, die sie den abgefragten Positionen nicht zuordnen konnten. Diese Textangaben werden während des Aufbereitungsprozesses analysiert und mit den sonstigen Einkommen des Haushalts abgeglichen. Falls es sich um zusätzliches Einkommen handelt, das an keiner anderen Position angegeben wurde, werden die Werte dieser Einkommensart den entsprechenden Einkommenskomponenten zugeordnet und somit ebenfalls in die Ermittlung der Einkommenssituation einbezogen. Die Erhebungsunterlagen enden jeweils mit einem Bemerkungsfeld, in dem die Befragten noch Anmerkungen zu den Fragen vornehmen können.

Plausibilisierung der erfassten Daten

Die Datenaufbereitung umfasst alle Arbeitsschritte von der Eingangskontrolle und der Datenerfassung über die Prüfung

der erfassten Daten, die Imputation von fehlenden oder fehlerhaften Werten bis hin zur Generierung der Zielvariablen. Um die hohe Qualität der Einkommensdaten zu gewährleisten, werden zwei Arten von Qualitätskontrollen in diesem Datenaufbereitungsprozess unterschieden:

1. laufende und
2. abschließende Qualitätskontrolle.

Die laufende Kontrolle der Datenqualität erfolgt während der Aufbereitung der Daten, die abschließende Kontrolle nach der Generierung der Zielvariablen (siehe Schaubild 1 auf Seite 142). Zur laufenden Qualitätskontrolle im Rahmen der Datenaufbereitung gehört eine mehrstufige Plausibilitätsprüfung.

Die erste Stufe der laufenden Plausibilitätskontrolle ist bereits in den Prozess der Datenerfassung in den Statistischen Landesämtern integriert. Die Statistischen Ämter der Länder führen die Feldphase durch beziehungsweise versenden die Erhebungsunterlagen, prüfen die zurückgesandten Erhebungsunterlagen (Rücklaufkontrolle)⁷ und erfassen schließlich die Angaben aus den Erhebungsunterlagen. Bereits während der Rücklaufkontrolle in den Statistischen Ämtern der Länder wird geprüft, ob der Haushalt unter anderem die Einkommensfragen beantwortet hat und ob diese Angaben plausibel sind. Ist dies nicht der Fall, kontaktiert das Statistische Landesamt den Haushalt telefonisch oder schriftlich und bittet um die entsprechenden Auskünfte. Des Weiteren werden während der Datenerfassung verschiedene dialoggestützte Plausibilitätsprüfungen durchgeführt. Die Prüfungen basieren dabei zum einen auf den aus dem aktuellen Sozialrecht bekannten Unter- und Obergrenzen beziehungsweise auf möglichen Haushalts- und Erwerbskonstellationen sowie aus den langjährigen Erfahrungen bei der Aufbereitung der Einkommens- und Verbrauchsstichprobe. Das Erfassungsprogramm enthält ein Bemerkungsfeld, in dem die Bearbeiter/-innen in den Statistischen Landesämtern zusätzlich Hinweise über die durchgeführten Rückfragen oder über Besonderheiten des Haushalts eintragen können.

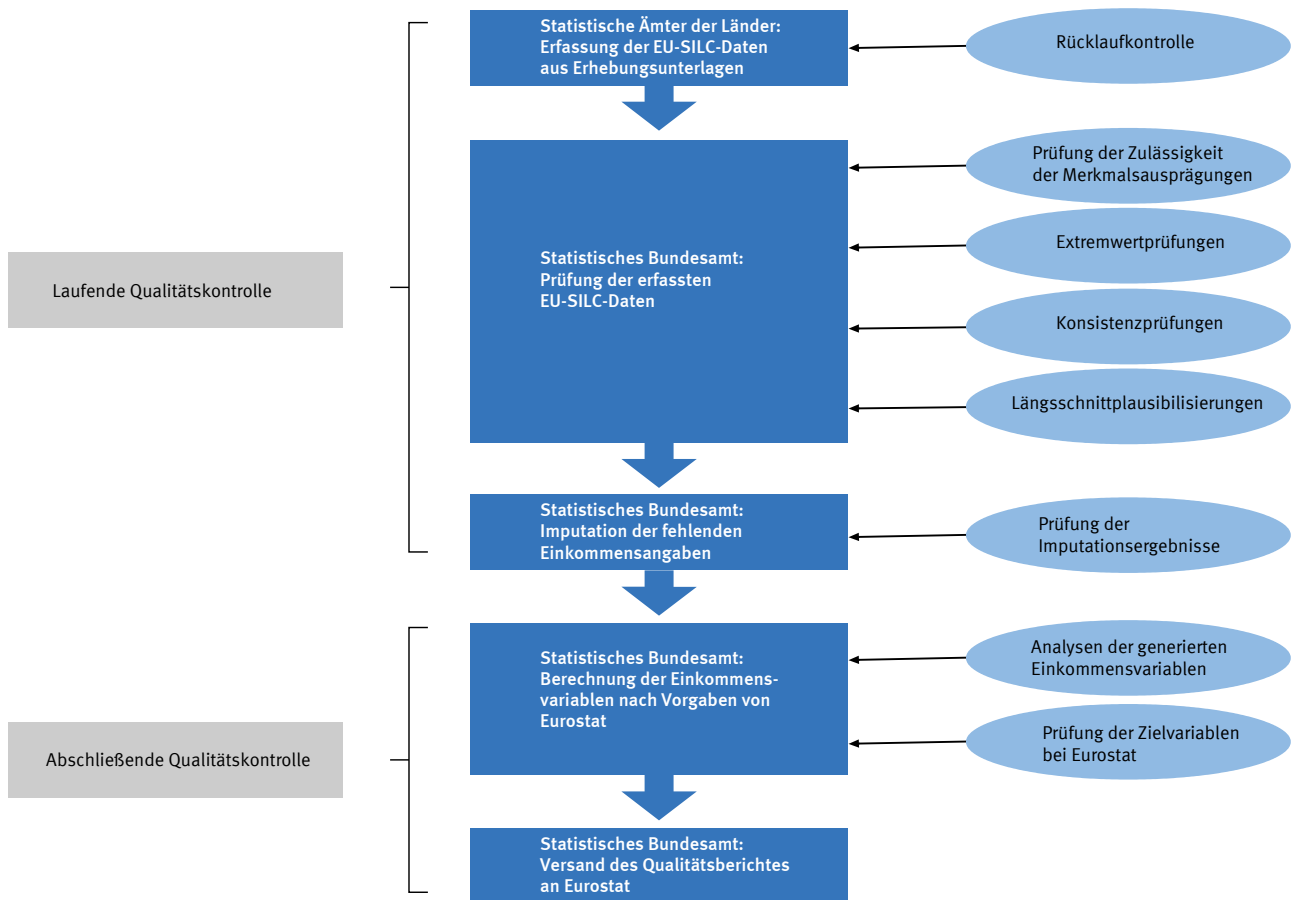
Die so erfassten Daten werden anschließend – ohne Namens- und Adressangaben – über ein speziell für die Übertragung von statistischen Einzelangaben eingerichtetes System an das Statistische Bundesamt gesandt. Das Statistische Bundesamt erstellt das Bundesergebnis, generiert die Zielvariablen nach den Vorgaben von Eurostat und liefert die endgültigen EU-SILC-Mikrodaten an Eurostat.⁸ Im Statistischen Bundesamt findet nach dem Zusammenspielen der länderspezifischen EU-SILC-Daten die zweite Stufe der laufenden Plausibilitätskontrolle statt. Der Dateninhalt wird in den folgenden vier Arbeitsschritten untersucht:

1. Prüfung der Zulässigkeit der Merkmalsausprägungen,
2. Extremwertprüfungen bei den Einkommensvariablen,

⁷ So wird zum Beispiel geprüft, ob es im Haushalt Veränderungen gab, die vor Beginn der Erfassung in die EU-SILC-Panelndatenbank eingepflegt werden müssen und/oder die den Versand eines weiteren Personenfragebogens an ein neues Haushaltsmitglied erfordern.

⁸ Siehe Sikorski, U.: „Europäische Panelerhebung über Einkommen und Lebensbedingungen“ in WiSta 8/2009, Seite 795 ff.

Schaubild 1 Mehrstufige Prüfung der EU-SILC-Daten



2012 - 01 - 0104

3. logische Prüfungen der Einkommensvariablen (Konsistenzprüfungen),
4. Längsschnittplausibilisierungen – Vergleich der Einkommensangaben einer Person beziehungsweise eines Haushaltes mit den Angaben aus der vorangegangenen Erhebung.

Prüfung der Zulässigkeit der Merkmalsausprägungen

Nach der dialoggestützten Plausibilitätsprüfung während der Datenerfassung werden im Statistischen Bundesamt zusätzliche ausführliche Einzelfallprüfungen durchgeführt. Bei den Einkommensdaten wird nicht nur die Zulässigkeit der Einkommenswerte selbst geprüft, sondern auch die Signierung der fehlenden Werte. Bezieht eine Person beziehungsweise ein Haushalt kein Einkommen einer bestimmten Art oder keinen Bestandteil einer bestimmten Einkommensart (zum Beispiel Privatentnahmen aus dem Betriebs-/Geschäftsvermögen beim Einkommen aus selbstständiger Tätigkeit oder Sozialgeld als Teilkomponente der sonstigen Leistungen gegen soziale Ausgrenzung und somit als Bestandteil des Einkommens aus empfangenen Sozialtransfers⁹), muss durch eine entsprechende Signierung darauf

⁹ Weitere Informationen zu Bestandteilen der Einkommensvariablen siehe Übersicht 2.

hingewiesen werden. Dadurch wird sichergestellt, dass die Daten weiter vollständig und korrekt aufbereitet werden.

Extremwertprüfungen bei den Einkommensvariablen

Bei Extremwerten handelt es sich um auffallend hohe oder niedrige Beobachtungswerte. Solche Werte sind besonders sorgfältig zu prüfen, weil es sich dabei auch um Eingabe- oder Angabefehler handeln kann, beispielsweise wenn der Befragte das monatliche Einkommen statt des Jahreseinkommens angegeben hat.

Um solche auffälligen Angaben zu identifizieren, werden zwei unterschiedliche Extremwertprüfungen durchgeführt. Zuerst werden für die Einkommensvariablen die erfassten Daten der Höhe nach sortiert und alle Werte ermittelt, die den Wert der untersten 0,1% der beobachteten Werte unterschreiten oder den Wert der obersten 0,1% der beobachteten Werte überschreiten. Anschließend wird durch Einzelfallbetrachtungen untersucht, ob die jeweiligen Angaben korrekt sind.

Im zweiten Schritt der Extremwertprüfung werden jene Beobachtungswerte überprüft, die außerhalb eines bestimmten Wertebereichs liegen (in der Regel bildet der Bereich die Werte zwischen dem oberen und unteren Quartil ab). Es wer-

den vor allem jene Beobachtungswerte näher untersucht, deren Abstand zu den Grenzen des oberen beziehungsweise unteren Quartils das 1,5-Fache beziehungsweise 2-Fache des Interquartilsabstands ($Q_{75} - Q_{25}$) übersteigt. Die auf diese Weise identifizierten Extremwerte werden den Statistischen Ämtern der Länder mitgeteilt, damit diese durch Rückfragen bei den Haushalten die Richtigkeit der Angaben kontrollieren.

Logische Prüfungen der Einkommensvariablen (Konsistenzprüfungen)

Die Prüfungen, ob die Einkommensangaben jeder Erhebungseinheit (Haushalt oder Person) konsistent sind, stellen den umfangreichsten Abschnitt der Datenprüfung dar. Dazu zählen die Einkommensangaben aus dem Personenfragebogen und aus dem Haushaltsfragebogen ebenso wie übergreifende Prüfungen der Personen- und Haushaltsangaben.

Die Personendaten zu den verschiedenen Bestandteilen des Einkommens aus abhängiger Tätigkeit, zum Einkommen aus selbstständiger Tätigkeit, zu den Renten, Pensionen und Leistungen öffentlicher Träger, zu den gezahlten Steuern und Sozialversicherungsbeiträgen und zu den Angaben zu privaten Vorsorgeformen werden besonders geprüft. Die Prüfungen umfassen auch die Lebenssituation der befragten Person im Einkommensreferenzjahr (das Jahr vor dem Erhebungsjahr), da unterstellt wird, dass die Lebenssituation stark mit den verschiedenen Einkommensarten korreliert. Bei der Plausibilisierung wird zum Beispiel geprüft, ob einem Monat Ruhestand nicht Monate der Vollzeitbeschäftigung folgen oder ob das Alter einer Zivildienst beziehungsweise Wehrdienst leistenden Person nicht über der Altersgrenze liegt. Des Weiteren werden spezifische Inkonsistenzen zwischen den Angaben zur Lebenssituation und zur Höhe des Einkommens kontrolliert (zum Beispiel dauerhaft erwerbsunfähige oder aus anderen Gründen nicht erwerbstätige Personen und deren Angaben zum Einkommen) sowie die Konsistenz unterschiedlicher Einkommensangaben. Bei der Konsistenzprüfung der Personenangaben werden insgesamt etwa 200 Prüfschritte durchgeführt.

Die Plausibilitätskontrollen der Einkommensangaben im Haushaltsfragebogen umfassen insgesamt knapp 100 Prüfschritte. Dabei werden die Angaben sowohl der Höhe nach auf ihre Plausibilität geprüft (zum Beispiel Angaben zum Betrag des Arbeitslosengeldes II und so weiter) als auch die Konsistenz der Angaben (zum Beispiel Monate des Bezugs des Arbeitslosengeldes II und der Grundsicherung oder Vergleich der Beträge für Zinsen und Tilgung bei den Hypothekenkrediten).

Anschließend werden einige einkommensrelevante Daten aus dem Haushaltsfragebogen mit Informationen aus dem Personenfragebogen übergreifend geprüft, insbesondere Angaben über die erhaltenen staatlichen Leistungen.

Längsschnittplausibilisierungen

Die EU-SILC-Erhebung ist sowohl als Querschnitt- als auch als Längsschnitterhebung konzipiert.¹⁰ Bei einer Längs-

schnittplausibilisierung werden die Daten aus der aktuellen Erhebung mit den Daten aus der letzten (vorangegangenen) Erhebung verglichen. Da die EU-SILC-Erhebung als eine rotierende Panelerhebung¹¹ angelegt ist, bezieht sich diese Prüfung auf drei Viertel der Haushalte in der Stichprobe. Für das im jeweiligen Erhebungsjahr neu hinzugekommene Viertel der Haushalte kann erst im Folgejahr eine Längsschnittplausibilisierung erfolgen.

Die Einkommensangaben aus der aktuellen Erhebung werden mit denen aus der vorangegangenen Erhebung verglichen. Für jeden Einkommensvergleich ist ein Fehlerindex eingebaut, mit dem der zulässige Abweichungsbereich für Änderungen jeder Einkommenskomponente bestimmt wird. Als Beispiele seien Renten der gesetzlichen Rentenversicherung und Einkommen aus selbstständiger Tätigkeit genannt: Während die Renten in der Regel eher nur kleinen Schwankungen unterliegen, können bei den Einkommen aus selbstständiger Tätigkeit größere Abweichungen auftreten.

Die Längsschnittplausibilisierungen dienen unter anderem auch dazu, Verwechslungen von Monats- und Jahresangaben aufzudecken. Auffallende Daten, die nicht auf fehlerhafte Eingaben während der Erfassung zurückzuführen sind, werden durch Rückfragen bei den Haushalten geklärt.

Zielsetzung und Anforderungen an die Einkommensimputation

Das grundsätzliche Ziel der Imputation von unvollständigen Einkommensdaten in einer Erhebung wie EU-SILC ist es, in der Erhebung vorhandene Informationen für die Generierung der fehlenden Werte zu nutzen. In der Regel erfolgen Imputationen vor allem dann, wenn grundsätzlich persönliche Angaben vorliegen und nur einzelne Einkommensfragen nicht beantwortet wurden. Hierbei handelt es sich um den sogenannten Item-Non-Response, das heißt um fehlende Informationen bei einzelnen Fragen von Haushalten/Personen, die an der EU-SILC-Erhebung teilgenommen haben.

Nicht alle beteiligten Haushalte und Personen beantworten die sensiblen und umfangreichen Fragen zum Einkommen bei EU-SILC vollständig. Trotz der Rückfragen durch die Statistischen Landesämter bei den Haushalten sind in diesem Bereich Antwortausfälle nicht zu vermeiden. Fehlende Angaben reduzieren aber nicht nur die statistische Basis für die Auswertungen, sondern können auch die Validität der Daten negativ beeinflussen. Für die wichtigsten Einkommensaggregate wie zum Beispiel das gesamte verfügbare Haushaltseinkommen würden die Antwortausfälle einen Verlust für die Aggregation notwendiger Einkommensbestandteile bedeuten.

Bei dem Vorgänger von EU-SILC, dem Europäischen Haushaltspanel (European Community Household Panel – ECHP, in den Jahren 1994 bis 2001 durchgeführt) erfolgte die Imputation der Einkommensangaben direkt bei Eurostat. Die für das Europäische Haushaltspanel angewandte zen-

¹⁰ Siehe Sikorski, U./Horneffer, B./Kuchler, B.: „Die Längsschnittdaten von EU-SILC – Datenstruktur und Hochrechnungsverfahren“ in WiSta 11/2009, Seite 1117 ff.

¹¹ Die EU-SILC-Stichprobe ist in vier Unterstichproben (Rotationsgruppen) unterteilt, deren Zusammensetzung jeweils repräsentativ für die Gesamtbevölkerung ist. Jedes Jahr scheidet eine Rotationsgruppe aus (Vierjahres-Rotationspanel).

trale Imputation der Einkommensdaten konnte aber nicht alle nationalen Gegebenheiten genau berücksichtigen. Aufgrund der vielen spezifischen steuer- und sozialpolitischen Rahmenbedingungen in den europäischen Ländern werden die Imputationen der Einkommensdaten bei EU-SILC ausschließlich durch die Mitgliedstaaten durchgeführt. Eine Durchführungsverordnung von Eurostat regelt jedoch die methodischen Mindestanforderungen an die Imputation, unter anderem dürfen die Ergebnisse der Imputationsverfahren die Variation und Korrelation der jeweiligen Variablen nicht beeinträchtigen. Außerdem sollen bei EU-SILC möglichst solche Imputationsverfahren verwendet werden, bei denen „Fehlerkomponenten“ in die imputierten Werte eingebaut sind (anstelle solcher, bei denen lediglich ein erwarteter Wert imputiert wird).¹²

Imputationen mithilfe von redundanten Informationen

Aus den erhobenen Daten werden Hilfsvariablen und Zwischenvariablen berechnet beziehungsweise logisch abgeleitet, die für die Imputation und für die Generierung der Zielvariablen benötigt werden. Die Hilfsvariablen werden auf Basis der soziodemografischen Angaben (wie zum Beispiel Alter, Geschlecht, Beruf) gebildet und werden herangezogen, um Imputationsklassen im Rahmen einer statistischen regressionsbasierten Modellierung zu erzeugen. Die Zwischenvariablen erfassen die einzelnen Einkommenskomponenten wie zum Beispiel Arbeitslosengeld II, Lohn/Gehalt oder Renten aus der gesetzlichen Rentenversicherung. Die Einkommensimputation wird in Deutschland auf der Ebene dieser Zwischenvariablen durchgeführt.

Die vorhandenen Informationen, die für die Imputation genutzt werden, finden sich in den drei Dimensionen der EU-SILC-Paneldaten:

1. Die erste Dimension wird durch die Menge der erhobenen Variablen gebildet, indem sich eine Variable aus einer oder mehreren anderen ableiten beziehungsweise durch sie erklären lässt. Das bedeutet, dass Informationen innerhalb eines Sets von Variablen mehrfach erhoben werden, sodass ein fehlender Wert aus anderen ihn erklärenden Variablen erzeugt werden kann.
2. In der zweiten Dimension wird nach der Ähnlichkeit der Fälle innerhalb einer Stichprobe gesucht. Analog zur Ähnlichkeit von Variablen basiert diese Vorgehensweise darauf, dass sich einzelne Erhebungseinheiten in der Stichprobe in Bezug auf die erhobenen Attribute sehr ähneln oder nahezu identisch sind.
3. Die dritte Dimension berücksichtigt den zeitlichen Aspekt eines Panels, denn je nach dessen Dauer sind die Variablen für jede Erhebungseinheit über mehrere Wellen hinweg verfügbar.

¹² Verordnung (EG) Nr. 1981/2003 der Kommission vom 21. Oktober 2003 zur Durchführung der Verordnung (EG) Nr. 1177/2003 des Europäischen Parlaments und des Rates für die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC) im Hinblick auf die Aspekte der Feldarbeit und die Imputationsverfahren (Amtsblatt der EU Nr. L 298, Seite 23).

Diese drei Dimensionen bilden die Grundlage dafür, die Redundanz von Informationen und die Wege, diese bei der Imputation zu nutzen, zu systematisieren.¹³ Für jede Dimension wird ein entsprechendes Imputationsverfahren eingesetzt.

Deduktive Imputationsverfahren

Die erste Gruppe der Imputationsmethoden, die in EU-SILC eingesetzt werden, umfasst die sogenannten deduktiven Verfahren. Diese Verfahren sind dadurch charakterisiert, dass die fehlenden Werte durch logische Ableitung von anderen Variablen oder basierend auf externen Quellen (zum Beispiel den Gesetzen des Sozialrechts) imputiert werden können.

Zu den deduktiven Imputationsverfahren gehört die Brutto-Netto-Konvertierungsmethode, das heißt die Umwandlung von Lohnabzügen, wie beispielsweise Sozialversicherungsbeiträgen, in Bruttobeträge unter Berücksichtigung der gesetzlichen Regelungen zu Steuern und Sozialabgaben. Dabei werden in einem iterativen Verfahren – ausgehend von einem initialen Bruttobetrag – so lange zum Beispiel Sozialversicherungsbeiträge berechnet und anschließend der der Berechnung zugrunde liegende Bruttobetrag entsprechend verändert, bis der Sozialversicherungsbeitrag aus dem Iterationsprozess mit dem erhobenen Betrag übereinstimmt. Angewendet wird diese Methode bei allen Einkommenskomponenten, für die neben den Bruttobeträgen auch Angaben zu Steuern oder Sozialabgaben erfragt werden, zum Beispiel bei Einkommen aus unselbstständiger Tätigkeit oder Altersleistungen.

In EU-SILC kann auf die jeweils angegebenen Beträge zu Steuern und Sozialversicherungen zurückgegriffen werden. Je nachdem, welche vom Bruttoeinkommen abgeleiteten Größen zur Verfügung stehen, lassen sich zwei Konvertierungsvarianten unterscheiden. Insbesondere für solche Einkommen aus abhängiger Beschäftigung, die die für geringfügig entlohnte Beschäftigung gesetzlich festgelegte Grenze (400 Euro monatlich) überschreiten, jedoch unter den Beitragsbemessungsgrenzen der Sozialversicherungen liegen (im Jahr 2009 waren das für die Kranken- und Pflegeversicherung 3 675 Euro monatlich), können die Sozialabgaben eine geeignete Grundlage für eine Konvertierung zurück zu den der Berechnung des Lohnabzugs zugrunde liegenden Bruttowerten bilden. Als Voraussetzung muss allerdings zuvor der eindeutige Bezug der angegebenen Steuern und Sozialversicherungsbeiträge zur gesuchten Einkommenskomponente sichergestellt sein. Dabei erfolgen bei den Sozialabgaben Plausibilitätsprüfungen, wie sie bereits in ähnlicher Weise für die Einkommensangaben verwendet wurden.¹⁴ Analog zur Konvertierung mithilfe der Sozialabgaben bilden für die zweite Konvertierungsvariante die angegebenen Steuern die Grundlage. Wegen der steuerlichen Grundfreibeträge eignet sich dieses Verfahren nicht

¹³ Zur Redundanz von Paneldaten siehe Minkel, H.: „Methodik zur Konvertierung von Paneldaten unter Ausnutzung zusätzlicher Informationen aus anderen Erhebungen“ in WiSta, Sonderausgabe ISI-Weltkongress 2003, Seite 76 f., hier: Seite 77.

¹⁴ Siehe dazu Gauckler, B./Körner, T./Minkel, H.: „Arbeitgeber sozialbeiträge und Beiträge zur Altersvorsorge“ in WiSta 3/2007, Seite 247 ff.

für niedrige Einkommen, lässt sich insbesondere bei höheren Einkommen aber im Gegensatz zu dem erstgenannten sehr gut einsetzen. Für den Fall, dass sowohl Sozialabgaben als auch Steuern angegeben wurden, können beide Konvertierungsverfahren angewandt werden. Als Bruttobetrag wird dann der höhere Wert aus den beiden Verfahren ausgewählt. Durch diese Vorgehensweise wird das Ausmaß einer möglichen Unterschätzung des Bruttoeinkommens reduziert.

Um die steuer- und sozialabgabenrechtlichen Regelungen möglichst umfassend in den Konvertierungsverfahren berücksichtigen zu können, gehen in die Modellierung zahlreiche Variable mit ein. Für die Lohnsteuerberechnung ist von zentraler Bedeutung, welche Steuerklasse dem/der Befragten zugeordnet wird. Für eine möglichst realitätsnahe Zuordnung wird neben dem Familienstand und der steuerrechtlichen Definition von Alleinerziehenden auch die Relation zwischen dem eigenen Einkommen des/der Befragten und dem der Ehepartnerin beziehungsweise des Ehepartners berücksichtigt. Ebenfalls wichtig für die Berechnung der Steuern ist die Frage, inwieweit in diesen auch Kirchensteuern enthalten waren. Da jedoch zur Religions- oder Kirchenzugehörigkeit keine Informationen im Fragebogen vorhanden sind, muss eine entsprechende Hilfsvariable auf Grundlage soziodemografischer Angaben modelliert werden. Eine statistische Zuordnung der Zugehörigkeit zur evangelischen oder römisch-katholischen Kirche erfolgt daher über das Produkt der Randwahrscheinlichkeiten einer entsprechenden Kirchenzugehörigkeit nach Bundesland, Geschlecht, Schulabschluss und Alter (die letzten drei Merkmale differenziert nach alten und neuen Bundesländern) auf Grundlage von Datenblättern der Forschungsgruppe Weltanschauungen in Deutschland¹⁵.

Zu den eingesetzten deduktiven Imputationsmethoden zählen neben den genannten Konvertierungsverfahren auch solche Verfahren, bei denen aufgrund sozialrechtlicher Bestimmungen und durch zum Teil sehr umfangreiches Einbeziehen der dafür notwendigen Zwischenvariablen Sozialtransfers, wie Kindergeld, Unterhaltsvorschuss oder Wohngeld, modelliert werden.

Statistische Imputationsverfahren

Zu einer zweiten Gruppe von Imputationsmethoden in EU-SILC lassen sich all jene Verfahren zusammenfassen, die auf personen- oder haushaltsübergreifenden statistischen Analysen basieren und so die Ähnlichkeit von Fällen innerhalb der Stichprobe berücksichtigen. Innerhalb von je nach Fragestellung in unterschiedlicher Weise zusammengefassten Gruppen ähnlicher Fälle werden die Zusammenhänge zwischen zu erklärenden (abhängigen) und erklärenden (unabhängigen) Variablen für jede der gebildeten Gruppen durch Regressionsanalysen genutzt. Insbesondere bei der Imputation wichtiger Einkommenskomponenten wie den Einkommen aus selbstständiger oder unselbstständiger

Tätigkeit wird auf komplexe Modellierungen zurückgegriffen. Dabei werden für jeden zu imputierenden Einkommenswert verschiedene lineare Regressionsmodelle berechnet und für die Imputation das Regressionsmodell ausgewählt, das das höchste Bestimmtheitsmaß¹⁶ aufweist. Die Regressionsmodelle unterscheiden sich durch die Anzahl der Variablen beziehungsweise durch die Anzahl der jeweiligen Ausprägungen der Variablen, die für die Bildung von Imputationsklassen verwendet werden. Imputationsklassen sind Merkmalsgruppen, in welche die Stichprobe eines vollständig erhobenen Merkmals eingeteilt werden kann. Diese werden so definiert, dass die Stichprobenwerte innerhalb der Klasse möglichst homogen in Bezug auf die zu schätzenden Einkommenswerte sind und sich von den Stichprobenwerten der anderen Klassen unterscheiden. Jede innerhalb der Stichprobe auftretende Kombination von Merkmalsausprägungen bildet dann eine Imputationsklasse für die Imputation einer Einkommenskomponente. Innerhalb der Imputationsklasse wird die Regressionsfunktion ermittelt – vorausgesetzt, dass eine je nach Einkommenskomponente unterschiedlich vorgegebene Mindestfallzahl von Stichprobeneinheiten erreicht wird. Nicht dazu zählen Stichprobeneinheiten, deren Werte außerhalb eines vorgegebenen Bereichs liegen. Damit sollen unplausible Daten bei der Regressionsanalyse ausgeschlossen werden. Bei der kombinatorischen Vielfalt von Merkmalsausprägungen aller Variablen ist es nicht immer möglich, für alle zu imputierenden Fälle auch eine Imputationsklasse mit identischen Ausprägungen zu finden. Deshalb werden die Imputationsklassen nicht nur auf Grundlage aller Variablen gebildet, sondern auch für jede beliebige Kombination einer Teilmenge der Variablen, unter Weglassen von einer, dann zweier Variablen und so weiter bis hin zum Ausschluss der gesamten Variablen, sodass die Imputationsklasse aus allen Stichprobeneinheiten besteht. Für jede Variablenkombination wird ein Regressionsmodell berechnet.

Werden zum Beispiel vier Variablen z_1 mit m_1 , z_2 mit m_2 , z_3 mit m_3 , z_4 mit m_4 Ausprägungen für die Bildung der Imputationsklassen gewählt, so erhält man insgesamt 16 Modelle:

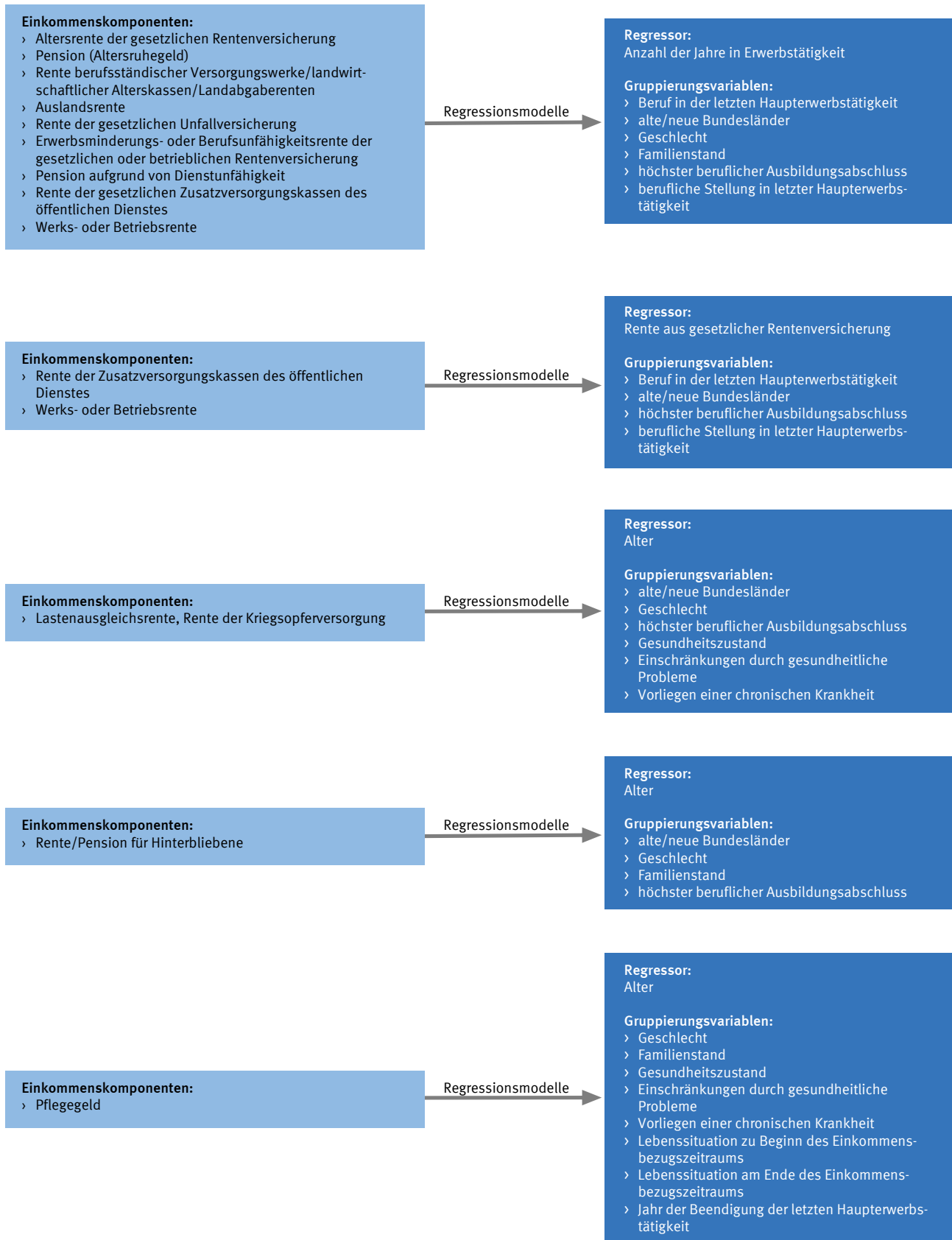
- > 1. Modell: $z_1 \times z_2 \times z_3 \times z_4$ mit $m_1 \cdot m_2 \cdot m_3 \cdot m_4$ Imputationsklassen
- > 2. bis 5. Modell: $z_i \times z_j \times z_k$ mit $m_i \cdot m_j \cdot m_k$ Imputationsklassen; $i, j, k = 1, \dots, 4$
- > 6. bis 12. Modell: $z_i \times z_j$ mit $m_i \cdot m_j$ Imputationsklassen; $i, j = 1, \dots, 4$
- > 13. bis 15. Modell: z_i mit m_i Imputationsklassen; $i = 1, \dots, 4$
- > 16. Modell: alle Stichprobeneinheiten

Insgesamt werden mit k Variablen 2^k verschiedene Regressionsmodelle erstellt. Für jeden fehlenden Wert werden damit 2^k Regressionswerte berechnet und als Imputationswert

¹⁵ Die Forschungsgruppe Weltanschauungen in Deutschland (fowid) greift in ihrer Arbeit auf bereits vorhandene empirische Datensammlungen zurück. Quelle für die eigenen Berechnungen der Randwahrscheinlichkeiten war hier die ALLBUS-Studie 2004, Studien-Nr. 3762 (ALLBUS: Allgemeine Bevölkerungsumfrage der Sozialwissenschaften; siehe www.gesis.org/allbus/studienprofile/2004/).

¹⁶ Das Bestimmtheitsmaß ist der Anteil der Varianz der abhängigen Variablen, der durch die lineare Regression erklärt wird. Es liegt zwischen 0 (kein linearer Zusammenhang) und 1 (die abhängige Variable wird vollständig durch das lineare Regressionsmodell erklärt).

Schaubild 2 Regressionsmodelle in EU-SILC am Beispiel der Renten/Pensionen



der Wert des Modells mit dem höchsten Bestimmtheitsmaß gewählt. Mit dieser Vorgehensweise werden in der Regel die Modelle bevorzugt, die auf der Einteilung der Stichprobe in viele kleine Imputationsklassen beruhen. Um eine zufällige Erklärung des Schätzers durch die Bildung sehr kleiner Imputationsklassen zu vermeiden, wird eine Mindestfallzahl für die Imputationsklassen vorgegeben. Zusätzlich werden nur Regressionswerte berücksichtigt, die in einen vorgegebenen Bereich fallen.

Zum Beispiel werden für die Imputation des Einkommens aus abhängiger Beschäftigung die sieben Variablen Branche (NACE, fünfstellig), Beruf (ISCO vierstellig), alte/neue Bundesländer (Ost/West), Ausbildungsabschluss (15 Ausprägungen), in Ausbildung befindlich (ja/nein), berufliche Stellung und Führungsposition (ja/nein) verwendet. Aufgrund der Vielzahl von Merkmalsausprägungen bei den hierarchisch aufgebauten Klassifikationsvariablen Branche und Beruf gehen diese Variablen auch alternativ mit vergrößerten Ausprägungen (vierstellig, dreistellig, zweistellig, einstellig) in die Modellierung mit ein. Die Branche kann somit in fünf verschiedenen Klassifikationsstufen oder natürlich auch gar nicht berücksichtigt werden, das heißt es ergeben sich hier sechs Möglichkeiten, beim Beruf sind dies fünf, bei allen anderen nicht hierarchisch aufgebauten Hilfsvariablen zwei (berücksichtigt oder nicht berücksichtigt). Daraus resultieren dann mit $k = 6 \cdot 5 \cdot 2^5$ zusammen 960 unterschiedliche Variablenkombinationen beziehungsweise Modellierungsansätze mit jeweils einer Vielzahl von Ausprägungskombinationen. Jede der auftretenden Ausprägungskombinationen aus diesen 960 Modellierungsansätzen mit einer Auftretenshäufigkeit über der vorgegebenen Mindestfallzahl wird dann als Imputationsklasse definiert. Für jede Imputationsklasse werden Regressionskoeffizienten berechnet, wobei es sich im genannten Beispiel bei dem zu erklärenden Regressanden um den Quotienten von Bruttomonatslohn und geleisteten Wochenarbeitsstunden handelt. Als Regressor dient die Anzahl der Jahre in Erwerbstätigkeit. Alle in Regressand und in Regressor oder in gegebenenfalls gebildete Quotienten eingehenden Größen können mithilfe von Modellierungsparametern in Bezug auf ihre Ausprägungen mit Minimal- und Maximalwerten spezifisch für die einzelnen Einkommenskomponenten eingegrenzt werden. Das heißt nur solche Fälle werden bei der Regression berücksichtigt, die auch über die entsprechenden Ausprägungen innerhalb der vorgegebenen Grenzen bei den analysierten Größen verfügen.

Für jede Einkommensvariable kann das beschriebene Verfahren mehrfach angewandt werden, mit jeweils unterschiedlichen Regressoren und Gruppierungsvariablen. Die Auswahl von einzelnen Gruppierungsvariablen und unabhängigen Variablen für die verschiedenen Einkommenskomponenten am Beispiel der Renten/Pensionen ist in Schaubild 2 dargestellt.

Grundsätzlich lässt sich über einen zusätzlichen Modellierungsparameter auch eine beliebig skalierbare additive stochastische Komponente auf Grundlage der Standardabweichung der Residuen berücksichtigen. Dies hat sich jedoch in der Regel als nicht notwendig erwiesen, da allein durch die große Vielfalt der berücksichtigten Modelle in

beträchtlichem Umfang Varianz in den Imputationsprozess eingebracht wird. Das bedeutet, dass die Summe der erklärten Varianz aus den letztendlich eingesetzten Regressionsmodellen einerseits und der Varianz zwischen den auf unterschiedlichen Ausprägungskombinationen basierenden Imputationsklassen sowie der Vielzahl an Modellierungsansätzen andererseits für eine wie von der Verordnung geforderte Bewahrung der Varianz völlig ausreicht.

Den statistischen Imputationsverfahren können auch all jene Methoden zugerechnet werden, die auch die Informationen aus der vorhergehenden Welle einbeziehen. Bei den Einkommenskomponenten, deren Wert im Laufe des Jahres identisch oder fast identisch bleibt, wie zum Beispiel bei den Rentenbeträgen, wird erst versucht, den Wert der gleichen Erhebungseinheit aus der vorangegangenen Erhebung zu verwenden. Falls der Wert aus der vorangegangenen Welle nicht vorhanden ist, die Erhebungseinheit noch nicht an der EU-SILC-Erhebung teilgenommen hat oder bei den bestimmten Einkommensangaben nicht von einer absoluten Unveränderbarkeit des Merkmals ausgegangen werden kann, muss auch hier eine regressionsbasierte Modellierung erfolgen. So gehen dann zum Beispiel auch empirisch in der Stichprobe zu beobachtende Einkommenssprünge, etwa beim Selbstständigeinkommen im Rahmen der konjunkturellen Veränderungen, in die Modellierung ein und werden bei der Imputation nachvollzogen.

Die Prüfung der Imputationsergebnisse bei EU-SILC erfolgt, indem die Parameter arithmetisches Mittel, Variationskoeffizient und Standardabweichung über alle Werte mit und ohne Imputationen verglichen werden (siehe das Beispiel in Tabelle 1).

Tabelle 1 Prüfung der Imputationsergebnisse nach dem Imputationsumfang

	Bruttoeinkommen aus unselbstständiger Tätigkeit in Form von Geld- oder geldwerten Leistungen				
	Imputations- ergebnisse		Arithme- tisches Mittel	Stan- dard- abwei- chung	Vari- ations- koeffi- zient
	Anzahl	%	EUR		%
vollimputiert	119	0,96	25 812	21 499	83,29
teilimputiert/ plausibilisiert . . .	75	0,6	19 999	18 100	90,50
downsize imputiert/ plausibilisiert . . .	406	3,27	65 200	60 232	92,38
nicht imputiert/ nicht plausibili- siert	11 833	95,17	27 353	23 415	85,60
Insgesamt . . .	12 433	100	28 530	26 307	92,21

Berechnung der Einkommensvariablen

Nachdem die fehlenden Einkommensangaben bei der Imputation vervollständigt wurden, werden die von Eurostat definierten Zielvariablen generiert. Die Einkommenszielvariablen bestehen häufig aus mehreren Einkommenskomponenten (Aggregation). In der Übersicht 2 ist am Beispiel der Erhebung 2010 dargestellt, wie die Einkommenskomponenten zugeordnet werden.

Übersicht 2

Zuordnung der Einkommenskomponenten

	Zielvariablen	Zugeordnete Einkommenskomponente
1	Einkommen aus unselbstständiger Tätigkeit	
1.1	Bruttoeinkommen aus unselbstständiger Tätigkeit in Form von Geld- oder geldwerten Leistungen	Lohn/Gehalt als Arbeitnehmer (Haupterwerb)
1.2	Bruttoeinkommen aus unselbstständiger Tätigkeit in Form von Sachleistungen	Lohn/Gehalt als Arbeitnehmer (Nebenerwerb) Private Nutzung des Firmenwagens Kostenlose oder ermäßigte Verpflegung (zum Beispiel Kantinenessen oder Essensgutscheine) Vom Arbeitgeber übernommene private Telefongebühren für ein dienstliches Mobiltelefon Sonstige Sach- und Naturalleistungen oder Rabatte
1.3	Firmenwagen	Private Nutzung des Firmenwagens
2	Einkommen aus selbstständiger Tätigkeit	
2.1	Bruttogewinn oder -verlust aus selbstständiger Tätigkeit	Bruttogewinn aus selbstständiger Tätigkeit Bruttoverlust aus selbstständiger Tätigkeit Privatentnahmen aus dem Betriebs-/Geschäftsvermögen
2.2	Wert der für den Eigenverbrauch produzierten Waren	Wert der im eigenen Garten oder mit eigener Kleintierhaltung produzierten Nahrungsmittel
3	Einkommen aus Vermögen	
3.1	Zinsen, Dividenden, Gewinne aus Kapitalanlagen in Unternehmen ohne eigene Rechtspersönlichkeit	Einnahmen aus Zinsen, Dividenden und Gewinnen (vor Abzug der Steuern) aus den Wertanlagen
3.2	Einkommen aus Vermietung oder Verpachtung	Bruttoeinkünfte (Pacht/Miete vor Abzug von Steuern, ohne Betriebskosten) Ausgaben für Instandhaltungskosten Ausgaben für Kreditzinsen (ohne Tilgung)
3.3	Rente aus privaten Rentensystemen	Rente aus einer privaten Vorsorge (Lebens-, Renten-, Berufsunfähigkeits- oder Unfallrentenversicherung) Leistungen aus privater Pflegezusatz- oder Krankentagegeldversicherung
4	Empfangene laufende Transfers	
4.1	Sozialleistungen	
4.1.1	Familienleistungen/Kindergeld	Mutterschaftsgeld Elterngeld Kindergeld Kinderzuschlag Unterhaltsvorschuss
4.1.2	Wohnungsbeihilfen	Wohngeld in Form von Miet- oder Lastenzuschuss Kosten der Unterkunft im Rahmen des Bezugs von Arbeitslosengeld II Kosten der Unterkunft im Rahmen des Bezugs von Grundsicherung
4.1.3	Arbeitslosenunterstützung	Arbeitslosengeld I Arbeitslosengeld II Kurzarbeitergeld, Schlechtwettergeld, Wintergeld, Konkursausfallgeld, Umschulungsgeld und Ähnliches Vorruhestandsgeld Abfindungen vor Erreichen des normalen Rentenalters beziehungsweise bei betriebsbedingten Kündigungen Förderung der Existenzgründung (Gründungszuschuss, Einstiegsgeld) Umschulungszuschüsse
4.1.4	Altersleistungen	Altersrente der gesetzlichen Rentenversicherung Pension (Altersruhegehalt) Rente der Zusatzversorgungskassen des öffentlichen Dienstes Werks- beziehungsweise Betriebsrente Rente berufsständischer Versorgungswerke/landwirtschaftlicher Alterskassen/Landabgaberenten Lastenausgleichsrente, Rente der Kriegsopferversorgung (für Personen ab 65 Jahren) Auslandsrente Rente der gesetzlichen Unfallversicherung (für Personen ab 65 Jahren) Erwerbsminderungs- oder Berufsunfähigkeitsrente der gesetzlichen oder betrieblichen Rentenversicherung (für Personen ab 65 Jahren) Pension aufgrund von Dienstunfähigkeit (für Personen ab 65 Jahren) Rente/Pension für Hinterbliebene (für Personen ab 65 Jahren) Abfindungen beim Eintritt in den Ruhestand Abfindungen bei betriebsbedingten Kündigungen (für Personen ab 65 Jahren) Pflegegeld (für Personen ab 65 Jahren)

noch Übersicht 2

Zuordnung der Einkommenskomponenten

	Zielvariablen	Zugeordnete Einkommenskomponente
4.1.5	Hinterbliebenenleistungen	Rente/Pension für Hinterbliebene (für Personen unter 65 Jahren)
4.1.6	Krankengeld	Krankengeld der gesetzlichen Krankenversicherung
4.1.7	Invaliditätsleistungen	Erwerbsminderungs- oder Berufsunfähigkeitsrente der gesetzlichen oder betrieblichen Rentenversicherung (für Personen unter 65 Jahren) Pension aufgrund von Dienstunfähigkeit (für Personen unter 65 Jahren) Rente der gesetzlichen Unfallversicherung (für Personen unter 65 Jahren) Lastenausgleichsrente, Rente der Kriegsopferversorgung (für Personen unter 65 Jahren) Pflegegeld (für Personen unter 65 Jahren)
4.1.8	Ausbildungsbezogene Leistungen	BAföG, Stipendium, Berufsausbildungsbeihilfe Zuschuss zur Krankenversicherung, Pflegeversicherung für BAföG-Studierende
4.1.9	Sonstige Leistungen gegen soziale Ausgrenzung	Sozialgeld (nach SGB II, ohne Arbeitslosengeld I oder II) Sozialhilfe (laufende Hilfe zum Lebensunterhalt, Leistungen nach dem 5. bis 9. Kapitel SGB XII) Bedarfsorientierte Grundsicherung im Alter und bei Erwerbsminderung Geldleistungen von Wohlfahrtsorganisationen (zum Beispiel Arbeiterwohlfahrt, Caritas, Diakonie)
4.2	Regelmäßig empfangene Geldtransfers zwischen privaten Haushalten	
4.2.1	Regelmäßig erhaltene Geldtransfers zwischen privaten Haushalten	Erhaltene Unterhaltszahlungen
4.2.2	Erhaltene Unterhaltszahlungen	Erhaltene sonstige Zahlungen Erhaltene Unterhaltszahlungen
5	Sonstiges Einkommen	
5.1	Von Personen unter 16 Jahren bezogenes Einkommen	Monatsnettoeinkommen und Anzahl der Monate des Einkommensbezugs bei Personen unter 16 Jahren
6	Geleistete Zinsen	
6.1	Zinsen für Hypothekenkredite	Zinsen für Hypothekenkredite (Betrag)
7	Geleistete laufende Transfers	
7.1	Einkommensteuern und Sozialbeiträge	Gezahlte Steuern (Personenfragebogen) Gezahlte Sozialversicherungsbeiträge (Personenfragebogen) Einkommensteuererstattung (Haushaltsfragebogen) Einkommensteuernachzahlung (Haushaltsfragebogen)
7.2	Regelmäßige Vermögensteuern	Grundsteuer für Grundbesitz
7.3	Regelmäßig geleistete Geldtransfers zwischen privaten Haushalten	Geleistete finanzielle Unterstützung an Personen außerhalb des Haushalts

Sobald alle personen- und haushaltsbezogenen Einkommensvariablen generiert sind, wird das Haushaltsbruttoeinkommen berechnet. Die wichtigste Größe für die Berechnung der Indikatoren über Armut und soziale Ausgrenzung ist das verfügbare Haushaltseinkommen. Das verfügbare Haushaltseinkommen ist definiert als das Haushaltsbruttoeinkommen abzüglich der Einkommensteuern und Sozialversicherungsbeiträge, der regelmäßigen Vermögensteuern und der regelmäßig geleisteten Geldtransfers zwischen privaten Haushalten (siehe Übersicht 1).

Neben der Generierung der Einkommensvariablen wird zu jeder Einkommensvariablen eine sogenannte Flag-Variable gebildet. Die Flag-Variablen haben die gleichen Namen wie die Einkommensvariablen selbst, ergänzt jedoch mit „_F“ am Ende. In den Flag-Variablen sind vielfältige Informationen über die Bildung der Einkommensvariablen kodiert.

Bezieht ein Haushalt/eine Person kein bestimmtes Einkommen, das heißt die jeweilige Einkommenszielvariable ist „0“, wird die Flag-Variable entsprechend mit „0“ kodiert.

Wenn ein Haushalt/eine Person ein bestimmtes Einkommen bezogen hat, wird die Information über die Bildung dieser Einkommenszielvariablen nach Vorgaben von Euro-

stat in der Flag-Variablen speziell gekennzeichnet. Aus dieser Kodierung kann geschlossen werden, ob es sich bei den erfassten Werten einer Variablen um Brutto- oder Nettowerte handelt, nach welcher Imputationsmethode (statische, Brutto-Netto-Konvertierung oder sonstige deduktive Imputationen) und in welchem Umfang imputiert wurde. Der Umfang der Imputation wird mithilfe eines Imputationsfaktors beziehungsweise einer Imputationskennziffer dargestellt. Der Imputationsfaktor wird als Quotient des erfassten (erhobenen) und des bei der Imputation und Plausibilisierung errechneten Wertes wie folgt berechnet und in Prozent (gerundet) ausgedrückt:

$$\text{Imputationsfaktor} = \text{erfasster Wert} \div \text{berechneter Wert} \cdot 100$$

Da es sich beim erfassten Wert einer Einkommensvariablen häufig um einen aggregierten Wert aus mehreren Einkommensbestandteilen handelt, weichen – wenn Einkommensbestandteile fehlten – der erfasste und der berechnete Wert voneinander ab.

Lag der erfasste Wert zum Beispiel bei 900 Euro und wurde bei der Imputation ein neuer Wert in Höhe von 1 100 Euro berechnet, beträgt der Imputationsfaktor 82 %. Der Impu-

tationsfaktor wird für jede Erhebungseinheit und für jede einzelne Einkommenszielvariable berechnet. Bei den in vollem Umfang imputierten Werten liegt der Imputationsfaktor bei 0. Wenn keine Imputationen für eine Zielvariable notwendig waren, hat der Faktor den Wert 100. Es kann auch vorkommen, dass der bei der Einkommensimputation berechnete Wert kleiner als der erfasste Wert ist. In diesem Fall wird nach den Vorgaben von Eurostat der Imputationsfaktor über 100 % liegen.

Der Imputationsumfang am Beispiel der ausgewählten Einkommensbestandteile empfangener laufender Transfers aus der EU-SILC-Erhebung 2010 ist in der Tabelle 2 dargestellt.

Tabelle 2 Imputationsumfang ausgewählter empfangener laufender Transfers in %

	Nicht imputiert	Teil-imputiert	Voll-imputiert
Familienleistungen/Kindergeld	99,25	0,72	0,03
Wohnungsbeihilfen	98,97	0,77	0,26
Arbeitslosenunterstützung	98,74	0,18	1,08
Altersleistungen	99,51	0,1	0,38
Hinterbliebenenleistungen	99,73	–	0,27
Ausbildungsbezogene Leistungen	100	–	–
Sonstige Leistungen gegen soziale Ausgrenzung	99,04	–	0,96
Regelmäßig erhaltene Geldtransfers zwischen privaten Haushalten	93,71	6,14	0,15
Erhaltene Unterhaltszahlungen	99,58	–	0,42

Inflationsfaktor bei Antwortausfällen im Haushalt

Der letzte Schritt bei der Generierung der Zielvariablen ist die Ermittlung des Inflations- oder Kompensationsfaktors für Haushalte mit Antwortverweigerungen einzelner Personen. Obwohl in Deutschland weniger als 1 % der Haushaltsmitglieder über 16 Jahre den Personenfragebogen nicht beantworten, können solche vollständigen Antwortausfälle von Personen (Partial Unit Non Response) grundsätzlich zu Verzerrungen der Ergebnisse und bei der Berechnung der Indikatoren zu Abweichungen führen. Aus diesem Grund wurde in den methodischen Vorgaben für EU-SILC ein haushaltsspezifischer Kompensationsfaktor für solche Antwortausfälle festgelegt. Dieser Faktor gibt an, mit welchem Wert das gesamte Haushaltseinkommen multipliziert (inflationiert) werden soll, um die Antwortausfälle zu kompensieren.

Haushalte, in denen alle Mitglieder ab 16 Jahren den Personenfragebogen beantwortet haben, erhalten hierbei den Wert 1. Bei den Haushalten mit Verweigerungen von Personen ab 16 Jahren (Partial Unit Non Response) werden auf Basis der regressionsbasierten Modellierungen Schätzwerte für das gesamte personenbezogene Einkommen (Löhne/Gehälter, bewertete private Nutzung des Firmenwagens, Gewinne oder Verluste der Selbstständigen, Arbeitslosenunterstützung, Altersleistungen, Hinterbliebenenrente, Krankengeld, Invaliditätsleistungen, ausbildungsbezogene Leistungen abzüglich Steuern und Sozialabgaben) der fehlenden Person berechnet. Dabei werden sowohl die Werte aus der

aktuellen Erhebung als auch die Vorjahresdaten als Regressoren berücksichtigt. Der Faktor wird als Quotient aus dem gesamten verfügbaren Haushaltseinkommen zuzüglich des imputierten Personeneinkommens und dem gesamten verfügbaren Haushaltseinkommen (ohne das imputierte Personeneinkommen) berechnet.

Prüfung der generierten Einkommensdaten bei Eurostat

Bevor die deutschen Daten an Eurostat verschickt werden, führt das Statistische Bundesamt verschiedene Analysen durch wie zum Beispiel:

- › Zeitreihenentwicklung bei den einzelnen Zielvariablen (gewichtet und ungewichtet),
- › Strukturvergleiche,
- › Analyse der Einkommensentwicklung insgesamt,
- › Analysen zur zeitlichen Entwicklung der Indikatoren,
- › Abschlussprüfung der EU-SILC-Mikrodaten auf der Grundlage von EU-weit standardisierten und jedem Mitgliedstaat zur Verfügung stehenden Programmen.

Nach der Lieferung der EU-SILC-Mikrodaten prüft Eurostat die Mikrodaten ebenfalls, und zwar in zwei Schritten:

1. Prüfung der Daten mit EU-weit standardisierten Programmen (EU-SILC checking programs),
2. interne Analysen von Eurostat.

Im ersten Schritt prüft Eurostat auf der Grundlage der sogenannten checking programs, ob die von dem Mitgliedstaat gelieferten EU-SILC-Mikrodaten die richtige Datenstruktur und die richtige Kodierung enthalten beziehungsweise ob nur die von Eurostat vorgegebenen Merkmalsausprägungen in den Daten vorhanden sind. Bei den Einkommensvariablen wird hier zusätzlich geprüft, ob die generierten Werte im erlaubten Wertebereich (zum Beispiel zwischen 0 und 99999999,99) liegen. Danach finden ausführliche Analysen der logischen Zusammenhänge zwischen bestimmten Einkommensinformationen statt, wie zum Beispiel Prüfungen, ob für jede Nettoeinkommensvariable eine entsprechende Bruttoeinkommensvariable vorliegt und umgekehrt oder ob die genannten Kinder unter 16 Jahren, die nach Angaben des Haushaltes eigene Einkünfte erzielen, tatsächlich zu diesem Haushalt gehören. Auch die Generierung des Brutto- und des verfügbaren Haushaltseinkommens wird untersucht. Dazu werden alle Einkommensbestandteile von allen Haushaltsmitgliedern in einer zusätzlichen Variablen aufsummiert und mit dem generierten Wert abgeglichen. Dabei finden zusätzliche Vergleiche des gesamten verfügbaren Haushaltseinkommens vor und nach Abzug von Sozialtransfers statt. Außerdem werden auch die Angaben zur Erwerbssituation im Zusammenhang mit den Einkommensinformationen kontrolliert. Im Fall von Längsschnittdaten findet zusätzlich eine Einzelfallprüfung bezüglich der Veränderungen in der Einkommenssituation (Vergleich mit

Vorjahresangaben) statt. Die bei den Prüfungen gefundenen auffälligen Werte werden den Mitgliedstaaten zur Korrektur oder Erklärung zugesandt.

Im zweiten Schritt führt Eurostat inhaltliche Analysen der Einkommensdaten durch; hierfür stehen den Mitgliedstaaten keine entsprechenden Programme zur Verfügung. Die Mitgliedstaaten erhalten lediglich einen Auszug mit Auffälligkeiten, die sie entweder korrigieren oder erklären müssen. Ziel dieser Prüfung ist, mögliche Ausreißer bei Haushalts- beziehungsweise Personeneinkommen zu identifizieren und die Entwicklung der Einkommen (zum Beispiel Median im zeitlichen Verlauf) zu untersuchen.

Erst nach erfolgreichem Abschluss dieser Datenkontrolle akzeptiert Eurostat die EU-SILC-Mikrodaten eines Mitgliedstaates. Anschließend können die Indikatoren zur Messung von Armut und sozialer Ausgrenzung veröffentlicht werden. In den von den Mitgliedstaaten jährlich zu erstellenden Qualitätsberichten sind unter anderem auch die eingesetzten Imputationsmethoden beschrieben und gegebenenfalls Abweichungen bei der Einkommensgenerierung von den methodischen Vorgaben für EU-SILC dokumentiert.

Die deutschen Querschnittsdaten mit den generierten Zielvariablen werden innerhalb von elf Monaten nach Beendigung der Feldphase, das heißt bis spätestens zum 1. Oktober des auf das Erhebungsjahr folgenden Jahres (Erhebungsjahr + 1), an Eurostat übermittelt. Für die Lieferung der Längsschnittdaten liegt der Termin innerhalb von fünfzehn Monaten nach dem Abschluss der Feldarbeit, das heißt bis zum 31. März des zweiten auf das Erhebungsjahr folgenden Jahres (Erhebungsjahr + 2).

Fazit

Eine der wichtigsten Aufgaben von EU-SILC ist es, eine fundierte Datenbasis für die europäische Sozialpolitik bereitzustellen. Aufgrund der Vielzahl von sensiblen Fragen zum Einkommen und der Komplexität der Einkommensfragen stellt eine detaillierte und vollständige Erfassung der Einkommen eine besondere Herausforderung für freiwillige Haushaltserhebungen dar. Seit Beginn der deutschen EU-SILC-Erhebung im Jahr 2005 wurde der Aufbereitungsprozess der Einkommensinformationen kontinuierlich und systematisch erweitert. Dabei müssen bei der Prüfung, Imputation und Generierung der Einkommensvariablen die methodischen Entwicklungen bei den einzelnen Zielvariablen, neu eingeführte Einkommenszielvariablen, jährliche Veränderungen bei den Sozialgesetzen und beim Steuerrecht berücksichtigt werden.

Auch bei Eurostat wurden die Datenprüfungen weiterentwickelt, um den hohen Qualitätsanforderungen gerecht zu werden. Die Implementierung des hier dokumentierten mehrstufigen Aufbereitungssystems bei der Generierung der Einkommensvariablen führt zu einer hohen Qualität der EU-SILC-Daten. [u](#)

Auszug aus Wirtschaft und Statistik

Herausgeber

Statistisches Bundesamt, Wiesbaden

www.destatis.de

Schriftleitung

Roderich Egeler, Präsident des Statistischen Bundesamtes

Brigitte Reimann (verantwortlich für den Inhalt)

Telefon: + 49 (0) 6 11 / 75 20 86

Ihr Kontakt zu uns

www.destatis.de/kontakt

Statistischer Informationsservice

Telefon: + 49 (0) 6 11 / 75 24 05

Telefax: + 49 (0) 6 11 / 75 33 30

Abkürzungen

WiSta	=	Wirtschaft und Statistik
MD	=	Monatsdurchschnitt
VjD	=	Vierteljahresdurchschnitt
HjD	=	Halbjahresdurchschnitt
JD	=	Jahresdurchschnitt
D	=	Durchschnitt (bei nicht addierfähigen Größen)
Vj	=	Vierteljahr
Hj	=	Halbjahr
a. n. g.	=	anderweitig nicht genannt
o. a. S.	=	ohne ausgeprägten Schwerpunkt
St	=	Stück
Mill.	=	Million
Mrd.	=	Milliarde

Zeichenerklärung

p	=	vorläufige Zahl
r	=	berichtigte Zahl
s	=	geschätzte Zahl
–	=	nichts vorhanden
0	=	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	=	Zahlenwert unbekannt oder geheim zu halten
...	=	Angabe fällt später an
X	=	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder —	=	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	=	keine Angaben, da Zahlenwert nicht sicher genug
()	=	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.