

Standard-Dokumentation Metainformationen

(Definitionen, Erläuterungen, Methoden, Qualität)

zu

EU-SILC 2003

Diese Dokumentation gilt für den Berichtszeitraum:
2002/2003

Diese Statistik war Gegenstand eines [Feedback-Gesprächs zur Qualität](#) am 02.06.2005

Bearbeitungsstand: **30.09.2005**



STATISTIK AUSTRIA
Bundesanstalt Statistik Österreich
A-1110 Wien, Guglgasse 13
Tel.: +43-1-71128-0
www.statistik.at

Inhaltsverzeichnis

1. Zusammenfassung, wichtige Hinweise	3
2. Allgemeine Informationen.....	4
Statistiktyp.....	4
Fachgebiet	4
Verantwortliche Organisationseinheit, Kontakt.....	4
Ziel und Zweck, Geschichte	4
Periodizität	5
Auftraggeber	5
Nutzer	5
Rechtsgrundlage(n).....	5
3. Statistische Konzepte, Methodik.....	5
Gegenstand der Statistik.....	5
Beobachtungs-/Erhebungs-/Darstellungseinheiten	6
Datenquellen.....	6
Meldeeinheit/Respondenten	6
Erhebungsform	6
Charakteristika der Stichprobe.....	6
Erhebungstechnik/Datenübermittlung.....	7
Erhebungsbogen (inkl. Erläuterungen).....	7
Teilnahme an der Erhebung.....	7
Erhebungs- und Darstellungsmerkmale, Maßzahlen; inkl. Definition	7
Verwendete Klassifikationen	9
Regionale Gliederung der Ergebnisse	9
4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen.....	9
Datenerfassung.....	9
Signierung (Codierung)	10
Plausibilitätsprüfung, Prüfung der verwendeten Datenquellen.....	10
Imputation (bei Antwortausfällen bzw. unvollständigen Datenbeständen)	10
Hochrechnung (Gewichtung)	12
Erstellung des Datenkörpers, (weitere) verwendete Rechenmodelle, statistische Schätzmethoden	12
Sonstige qualitätssichernde Maßnahmen	13
5. Publikation (Zugänglichkeit).....	14
Vorläufige Ergebnisse	14
Endgültige Ergebnisse	14
Revisionen	14
Publiziert in:	14
Behandlung vertraulicher Daten.....	14
6. Qualität	15
6.1. Relevanz.....	15
6.2. Genauigkeit.....	15
6.2.1. Stichprobenbedingte Effekte, Repräsentativität.....	15
6.2.2. Nicht-stichprobenbedingte Effekte	16
Qualität der verwendeten Datenquellen	16
Abdeckung (Fehlklassifikationen, Unter-/Übererfassung)	16
Antwortausfall (Unit-Non-Response, Item-Non-Response).....	17
Messfehler (Erfassungsfehler)	21
Aufarbeitungsfehler.....	21
Modellbedingte Effekte.....	22
6.3. Rechtzeitigkeit und Aktualität.....	22
6.4. Vergleichbarkeit	22
6.5. Kohärenz	23
Hinweis auf ergänzende Dokumentationen/Publikationen	32

1. Zusammenfassung, wichtige Hinweise

2003 fand in Österreich erstmals die EU-SILC-Erhebung (Gemeinschaftsstatistiken über Einkommen und Lebensbedingungen, *Community Statistics on Income and Living Conditions*) statt. 4600 Haushalte wurden über ihr Einkommen und weitere Merkmale zur sozialen Situation befragt. Die Feldarbeit fand von Juni bis Oktober statt und war an das Institut für empirische Sozialforschung (IFES) ausgelagert.

Mit „Einkommen“ ist das Netto-Einkommen des Haushalts im Jahr 2002 gemeint. Alle anderen Merkmale (Haushaltsgröße, Erwerbstätigkeit etc.) bezogen sich auf den aktuellen Status zum Zeitpunkt des Interviews.

EU-SILC befasst sich mit der Lebenssituation der Bewohner von privaten Haushalten. Die Teilnahme an der Erhebung ist freiwillig. Die Antwortquote lag im Jahr 2003 bei 64% – angesichts der Sensibilität des Haupterhebungsmerkmals Einkommen ein vergleichsweise hoher Wert.

EU-SILC ist eine europaweite Erhebung und bildet die Grundlage für vergleichende Einkommensstatistik sowie für Daten über Armut und Soziale Ausgrenzung. Von Eurostat, dem EU-Statistikamt, werden Zielvariablen vorgegeben, die von den Mitgliedstaaten für eine Mindestzahl von Haushalten geliefert werden müssen. Die konkrete Art der Datenerhebung bleibt den einzelnen Ländern überlassen. In Österreich fungierte auf nationaler Ebene das BMSG als Co-Financier.

EU-SILC 2003 war in Österreich eine reine Querschnitterhebung, d.h. sämtliche Respondenten scheiden nach diesem Jahr aus der Stichprobe aus. Ab 2004 beginnt das Panel, so bezeichnet man eine integrierte Quer- und Längsschnitterhebung.

Wie schon aus dem Europäische Haushaltspanel (ECHP) werden mit Inkrafttreten der Verordnung (2004) auch aus EU-SILC sowohl Quer- als auch Längsschnittindikatoren berechnet, allerdings mit verstärktem Augenmerk auf aktuellen und vergleichbaren Querschnittsdaten. Im Längsschnitt (ab 2004) wird eine Viertelrotation eingeführt, d.h. ein Haushalt wird in maximal vier aufeinander folgenden Jahren befragt.

In einer sozialwissenschaftlichen Studie wie EU-SILC sind v.a. Verteilungsfragen von Interesse, z.B. „Wie viel Prozent der Bevölkerung leben unter einem bestimmten Einkommenslevel?“ oder „Welche Bevölkerungsgruppen sind überdurchschnittlich in unteren Einkommenssegmenten vertreten?“. Dementsprechend wurden auch die statistischen Methoden an eine möglichst valide Messung dieses Zielkriteriums angepasst, etwa durch die Wahl dafür geeigneter Imputationsmodelle.

Eine Übersicht über die EU-SILC-Erhebungen ab 2002

Erhebungs-jahr	Art der Erhebung (Q = Querschnitt, L = Längsschnitt)	Rechtsgrundlage	Referenz-jahr Einkommen	Erscheinungsjahr der Publikation	Befragte Haushalte (netto)	Weiterverfolgung von Haushalten im Folgejahr? (j/n)
2002	Probeerhebung Q	Vertrag mit Eurostat	2001	(keine)	200	n
2003	Q	Verträge mit Eurostat und BMSG	2002	2005	4500	n
2004	Q, L integriert	EU-Verordnung	2003	2006	4500	j (3250 der 4500)
2005	Q, L integriert	EU-Verordnung	2004	2007	4500	j (3250 der 4500)
2006 ff.	Q, L integriert	EU-Verordnung	2005 ff.	2008 ff.	4500	j (3250 der 4500)

2. Allgemeine Informationen

Statistiktyp

Primärstatistische Erhebung.

Fachgebiet

Einkommen und Lebensbedingungen.

Verantwortliche Organisationseinheit, Kontakt

Bereich Soziales und Wohnen; Direktion Bevölkerung;

Mag. Martin Bauer

Tel. +43 (1) 71128-7021, E-Mail: martin.bauer@statistik.gv.at

Johannes Klotz

Tel. +43 (1) 71128-7749, E-Mail: johannes.klotz@statistik.gv.at

Ziel und Zweck, Geschichte

Ziel der Erhebung ist es, für sämtliche EU-Staaten vergleichbare Daten über die soziale Lage der Bevölkerung zu erhalten. EU-SILC soll auf europäischer Ebene die Grundlage für Statistiken über Einkommensverteilung, Armut und Soziale Eingliederung werden. Die Ergebnisse des Projekts fließen beispielsweise in die von der Kommission jährlich im Frühjahrsbericht veröffentlichten Strukturindikatoren und in die Laeken-Indikatoren ein.

Ab dem Beitritt zur Europäischen Union (1995) wurde auch in Österreich jährlich das Europäische Haushaltspanel (ECHP) durchgeführt. Diese Erhebung kann als Vorläuferin von EU-SILC angesehen werden und erfolgte 1994 bis 2001 auf Basis von Gentlemen's Agreements in allen EU-Staaten mit Ausnahme Schwedens. In Österreich wurde das ECHP vom Interdisziplinären Forschungszentrum Sozialwissenschaften (IFS) durchgeführt, wobei die Feldarbeit an eine Projektgemeinschaft aus Fessel-GfK und IFES ausgelagert war. Die siebente¹ und letzte Welle des ECHP fand 2001 statt. Zukünftig (ab 2004) werden die für Einkommensverteilung, Armut und soziale Ausgrenzung relevanten Informationen durch EU-SILC ermittelt werden.²

Nachdem es in Österreich nach dem Auslaufen des ECHP keine alternative Datenquelle gab, um für das Jahr 2002 die auf Europäischen Räten vereinbarten Indikatoren über den sozialen Zusammenhalt liefern zu können, musste im Jahr 2003 auf Basis bilateraler Verträge mit Eurostat und dem BMSG eine einmalige Querschnitterhebung durchgeführt werden.

EU-SILC unterscheidet sich vom ECHP in juristischer Hinsicht dadurch, dass es ab 2004 auf Basis einer EU-Verordnung, also einer für alle Mitgliedstaaten verbindlichen Rechtsgrundlage erfolgt (siehe w. u.). Weiters sind als methodische Differenzen insbesondere der vergrößerte Stichprobenumfang (in Österreich im Querschnitt mindestens 4500 Haushalte netto gegenüber zuletzt 2455 Haushalten im ECHP) und die Panelrotation (jedes Jahr Austausch von ca. ¼ aller Haushalte) zu nennen. Hinsichtlich der Befragungstechnik besteht für die Mitgliedstaaten größerer Freiraum, da es keinen EU-weit einheitlichen Fragebogen (*blueprint questionnaire*), sondern einen Zielvariablenkatalog gibt, dessen konkrete Umsetzung im Ermessen der Mitgliedstaaten liegt. Betreffend Österreich sind weiters die Umstellung von Papierfragebögen

¹ International gesehen handelt es sich um die achte Welle, da das ECHP in den damaligen EU-Mitgliedstaaten bereits 1994 gestartet wurde. Finnland begann 1996.

² Deutschland, Niederlande, Vereinigtes Königreich und die neuen Mitgliedsländer beginnen 2005.

auf CAPI-Interviews (*computer assisted personal interviewing*, Face-to-face-Interviews mit Laptop) und die Durchführung durch Statistik Austria zu nennen. Auffälligste inhaltliche Unterschiede sind zum einen die Verlagerung des Schwergewichts auf Querschnittsdaten, zum anderen die Erhebung von Bruttoeinkommen auf Haushaltsebene.

Ab der Erhebung 2004 sind zusätzliche Zielvariablen bezüglich Kinderbetreuung verbindlich. Auf Basis der Erhebung 2005 werden erstmalig Längsschnittdaten an Eurostat geliefert werden. Außerdem wird, ebenfalls ab 2005, jährlich ein Modul in die Befragung eingebaut werden, welches jeweils ein vertiefendes Thema zur sozialen Lage der Bevölkerung in Privathaushalten behandeln soll (*sekundäre Zielvariablen*).

Periodizität

Einmalige Querschnittserhebung. Ab 2004 wird auf Basis einer EU-Verordnung jährlich eine integrierte Quer- und Längsschnittserhebung durchgeführt, deren Querschnittsvariablenkatalog weitestgehend mit EU-SILC 2003 übereinstimmen wird.

Auftraggeber

Im Jahr 2003 Eurostat sowie das BMSG, jeweils auf Basis von bilateralen Verträgen mit Statistik Austria (siehe w. u.). Ab dem Erhebungsjahr 2004 ist die Republik Österreich auf Grund einer EU-Verordnung (VO 1177/2003 des Europäischen Parlaments und des Rates der EU) zur Durchführung von EU-SILC unmittelbar verpflichtet. Zur Finanzierung sind allerdings weiterhin Verträge notwendig.

Nutzer

- Europäische Union/Eurostat (v.a. zur Berechnung von Laeken- und Strukturindikatoren),
- BMSG (v.a. Sozialbericht),
- Armuts- und Sozialforschung an Universitäten, privaten Forschungsinstituten und bei Interessensvertretungen,
- Bundesländer (Landesstatistiker),
- Öffentlichkeit.

Rechtsgrundlage(n)

Bilaterale Verträge mit Eurostat (No. 200236400007) und dem BMSG (Werkvertrag GZ: 11.104/2-1/03). Es handelt sich also um ein einmaliges Auftragsprojekt.

Die EU-SILC-Erhebungen ab 2004 werden auf Basis einer Verordnung des Europäischen Parlaments und des Rates der EU sowie ausführender Kommissionsverordnungen erfolgen. EU-Verordnungen.

3. Statistische Konzepte, Methodik

Gegenstand der Statistik

Gegenstand der Statistik ist das Einkommen und die Lebensbedingungen der BewohnerInnen von Privathaushalten.

Beobachtungs-/Erhebungs-/Darstellungseinheiten

Dem Konzept nach umfasst der Erhebungsbereich von EU-SILC die Bevölkerung in Privathaushalten im Bundesgebiet. Nicht enthalten sind daher BewohnerInnen von Anstaltshaushalten bzw. Gemeinschaftsunterkünften sowie Personen ohne festen Wohnsitz. Ebenfalls nicht enthalten sind Grenzgänger aus dem Ausland (das sind Personen, die in Österreich arbeiten, aber ihren Hauptwohnsitz im Ausland haben, unabhängig von der Aufenthaltsdauer); Grenzgänger ins Ausland zählen dagegen zum Erhebungsbereich.

Erhebungseinheiten sind private Haushalte und deren BewohnerInnen. Der private Haushalt ist dabei ident mit den BewohnerInnen einer Wohnung. (Es ist daher nicht möglich, dass es innerhalb einer Wohnung mehrere Haushalte gibt.)

Aufbau der Datenfiles

Es werden vier Datenfiles generiert: Ein Haushaltsregister (D-File) für alle Auswahlhaushalte, eine Haushaltsdatei (H-File) für alle befragten Haushalte, ein Personenregister (R-File) für alle Personen in diesen Haushalten und eine Personendatei (P-File) für alle Personen ab 16 Jahren, mit denen ein Personeninterview durchgeführt wurde.³

Datenquellen

Keine, da Primärstatistik.

Meldeeinheit/Respondenten

Die Informationen des Haushaltsregisters waren durch die InterviewerInnen einzutragen. Für die Haushaltsdatei und das Personenregister war jeweils ein Haushaltsmitglied zu befragen, für die Personendatei die jeweilige Person. Fremdauskünfte (*Proxy-Interviews*) waren nur in Ausnahmefällen gestattet (Proxy-Quote: 14,4%). Sämtliche Auskunftspersonen mussten mindestens 16 Jahre alt sein.

Erhebungsform

Stichprobenerhebung.

Charakteristika der Stichprobe

Grundgesamtheit und Auswahlrahmen

Grundgesamtheit sind die privaten Haushalte, das sind Wohnungen mit zumindest einer Meldung mit Hauptwohnsitz.

Als Auswahlrahmen diente das Zentrale Melderegister (ZMR), Stand April 2003.

Samplingverfahren

8931 Adressen wurden wie folgt gezogen: Ein einstufiges Auswahlverfahren per Zufallsauswahl wurde in Wien, Vorarlberg und in den übrigen Bundesländern für Gemeinden ab einer bestimmten Größe angewandt.⁴ Die Adressen aus den restlichen Gemeinden wurden in einem zweistufigen Auswahlverfahren ermittelt, wobei zunächst die Gemeinden ausgewählt und anschließend aus jeder dieser Gemeinden 11 Adressen gezogen wurden (Klumpung). Nach Abschluss des Verfahrens bestand die Stichprobe aus 5892 Adressen aus 414 Gemeinden im

³ Referenzzeitpunkt war der 31.12.2002. Es handelt sich somit um Personen mit Geburtsjahr 1986 oder früher.

⁴ Als Kennzahl diente die Zahl der Wohnungen: mindestens 600 im Burgenland, 1000 in Salzburg, 1200 in Kärnten und Tirol, 1700 in der Steiermark und 2000 in Ober- und Niederösterreich.

einstufigen Bereich plus 3039 Adressen aus 293 Gemeinden im zweistufigen Bereich. (Zuvor waren noch einige Adressen ausgeschieden worden, bei denen es sich um falsch klassifizierte Anstaltshaushalte gehandelt hatte.)

Schichtung: im einstufigen Bereich nach Bundesland und Haushaltsgröße, im zweistufigen Bereich nach Bundesland und Agrarquote der Gemeinde.⁵

Auswahlwahrscheinlichkeit

Der Zahl der pro Schicht ausgewählten Adressen war direkt proportional zur Adressenzahl in der Grundgesamtheit, d.h. wenn auf eine Schicht 5% aller Hauptwohnsitzadressen Österreichs entfallen, so werden aus dieser Schicht auch 5% der Stichprobenadressen gezogen

Erhebungstechnik/Datenübermittlung

Die Feldarbeit war ausgelagert an das Institut für empirische Sozialforschung (IFES). Die Befragung wurde von 4. Juni bis 13. Oktober 2003 mittels CAPI (*computer assisted personal interviewing*) durchgeführt. Darunter versteht man *Face-to-face*-Interviews, bei denen die Antworten direkt in einen Tablet-PC eingegeben werden. Die Rohdaten wurden von den InterviewerInnen via Internet laufend an die IFES-Zentrale in Wien übermittelt und dort aufbereitet. IFES setzte ca. 105 InterviewerInnen ein, die im gesamten Bundesgebiet tätig waren.

Datenübermittlung an Statistik Austria

Entsprechend der Ausschreibung wurden dreimal Daten an Statistik Austria übermittelt: am 1.8., am 5.9. und am 3.11.2003, jeweils auf CD-ROM im ASCII-Format.

Erhebungsbogen (inkl. Erläuterungen)

Da es sich um eine CAPI-Erhebung handelte, sind Erhebungsbögen in Papierform insoweit vorhanden, als dass sie gemeinsam mit dem Codebook (Merkmalsverzeichnis) sowie den vorgegebenen [Plausibilitäts-Checks](#) die Grundlagen für die CAPI-Programmierung darstellten.

[Erhebungsunterlagen](#) (Fragenprogramm mit Erläuterungen, Karten und Listen).

Teilnahme an der Erhebung

Freiwillig.

Erhebungs- und Darstellungsmerkmale, Maßzahlen; inkl. Definition

Erhebungsmerkmale auf Haushaltsebene: Einkommenskomponenten auf Haushaltsebene (brutto und netto), Fragen zur Wohnung, zur Ausstattung mit Konsumgütern und zur finanziellen Lage, nicht-monetäre Deprivationsindikatoren.

Erhebungsmerkmale auf Personenebene: Einkommenskomponenten auf Personenebene (brutto und netto), demographische Daten, Fragen zur Erwerbstätigkeit, zur Bildung und zur Gesundheit.

Einkommenskomponenten in EU-SILC

Diese Komponenten werden sowohl brutto als auch netto erhoben. Das Nettoeinkommen ist das Bruttoeinkommen abzüglich der Beiträge zur gesetzlichen Sozialversicherung sowie der Einkommen- und Vermögensteuern. Fallen bei einer Komponente keine derartige Abzüge an (so wie in Österreich z.B. bei der Familienbeihilfe), so ist das Bruttoeinkommen gleich dem Nettoeinkommen.

Referenzzeitraum ist das dem Erhebungsjahr vorangegangene Kalenderjahr (2002).

⁵ Agrarquote bezeichnet den Anteil der land- und forstwirtschaftlichen Bevölkerung.

Komponenten auf Haushaltsebene

- Staatliche Beihilfen zu Wohnkosten
- Familienbeihilfe/Kinderabsetzbetrag
- Sonstige Familienleistungen (ausg. Wochengeld/Kinderbetreuungsgeld)
- Einkommen von Personen unter 16 Jahren

Komponenten auf Personenebene

- Einkommen aus unselbständiger Erwerbstätigkeit (inkl. privat genutzter Firmen-Pkw und anteilige Abfertigungen)
- Einkommen aus selbständiger Erwerbstätigkeit (inkl. Vermietung/Verpachtung)
- Vermögenseinkommen
- Einkommen aus Präsenz-/Zivildienst
- Arbeitslosenleistungen
- Pensionen und Unfallrenten
- Pflegegeld
- Krankengeld, Entgeltfortzahlung im Krankheitsfall, sonstige Krankheitsleistungen
- Staatliche Beihilfen bei Ausbildung (z.B. Stipendien)
- Sozialhilfe
- Wochengeld, Kinderbetreuungsgeld (inkl. Karenzgeld), Unterhaltsvorschüsse
- Sonstige staatliche Beihilfen
- Unterhaltszahlungen (erhalten wie geleistet)
- Sonstige private Transfers (erhalten wie geleistet)
- Steuernachzahlungen bzw. -rückerstattungen
- Sonstiges Einkommen

Die angeführten Komponenten wurden entweder auf Monats- oder auf Jahresbasis erfragt. Konnten oder wollten RespondentInnen einen Betrag nicht nennen, so wurden sie gebeten, auf einer vorgelegten Karte eine Stufe (Einkommensintervall) zu nennen.

Einen Sonderfall stellt das Einkommen aus Familienbeihilfe/Kinderabsetzbetrag dar. Nachdem die Sätze für diese Transfers bekannt sind, wurde diese Einkommenskomponente betragsmäßig nicht erhoben, sondern berechnet.

Das Haushaltseinkommen entspricht der Summe aller dieser Einkommenskomponenten im jeweiligen Haushalt. Ab 2007 wird das Haushaltseinkommen noch weitere Komponenten umfassen, z.B. unterstellte Mieten (*imputed rents*).

Darstellungsmerkmal Äquivalenzeinkommen

Die meisten EU-Indikatoren, die im Rahmen von EU-SILC erhoben werden, bauen auf dem Äquivalenzeinkommen auf. Darunter zu verstehen ist ein auf einen Einpersonenhaushalt standardisiertes Haushaltseinkommen. Verwendet werden die Gewichtungsfaktoren der EU-Skala (modifizierte OECD-Skala):

Erste erwachsene Person	1,0
Jede weitere Person ab 14 Jahren	0,5
Kinder von 0 bis 13 Jahren	0,3

Das Äquivalenzeinkommen wird berechnet, indem man das Haushaltseinkommen durch die Summe der Gewichte im Haushalt dividiert. Es fällt daher in Einpersonenhaushalten gleich hoch, ansonsten immer niedriger aus als das Haushaltseinkommen. Mit Hilfe dieser Größe sollen unterschiedlich große und verschieden strukturierte Haushalte vergleichbar gemacht werden.

Definition von Armutsgefährdung (Einkommensarmut)

Personen gelten, entsprechend der von EU/Eurostat verwendeten Definition, als armutsgefährdet, wenn sie in einem Haushalt leben, dessen äquivalisiertes Einkommen unterhalb eines Grenzwerts liegt. Dieser Grenzwert, die Armutsgefährdungsschwelle, wird bei 60% des medianen Äquivalenzeinkommens in einem Staat angenommen. Wenn also beispielsweise das mediane Äquivalenzeinkommen in einem Land bei 10.000 EUR jährlich liegt, so gelten jene Personen als armutsgefährdet, die in Haushalten leben, deren äquivalisiertes Einkommen niedriger als 6.000 EUR jährlich ist.

Es sei hier noch einmal darauf verwiesen, dass mit „Einkommen“ stets das Jahreseinkommen im dem Erhebungsjahr vorangegangenen Kalenderjahr gemeint ist, also bei der Erhebung 2003 das Einkommen im Jahr 2002.

Definition der Armutsgefährdungslücke

Darunter zu verstehen ist die prozentuelle Abweichung des medianen Einkommens armutsgefährdeter Haushalte von der Armutsgefährdungsschwelle.

Verwendete Klassifikationen

Die Adressen im Haushaltsregister waren gemäß EU-Vorgaben regional nach der Besiedlungsdichte⁶ sowie nach [NUTS 2](#) (in Österreich identisch mit den Bundesländern) zu klassifizieren.

In der Personenbefragung wurden, gemäß den Selbsteinschätzungen der RespondentInnen, die berufliche Tätigkeit nach der [ISCO-88-COM - Berufsklassifikation](#) (2-Steller) und der Wirtschaftszweig der Betriebsstätte bzw. Dienststelle nach der [ÖNACE 2003](#) - Systematik der Wirtschaftstätigkeiten (2-Steller) vercodet. Die Bildungsebene wurde analog zum Mikrozensus-Merkmal „höchste abgeschlossene Schulbildung“ erhoben und anschließend nach der [ISCED](#) (1-Steller) vercodet. Die berufliche Funktion war erhebungsseitig ebenfalls an ein Mikrozensus-Merkmal angelehnt, nämlich an die „Stellung im Beruf“.

Regionale Gliederung der Ergebnisse

Die auf Basis von EU-SILC 2003 berechneten Indikatoren werden grundsätzlich nicht regional gegliedert. In der Publikation wird eine Differenzierung nach Gemeindegrößenklassen verwendet.⁷ Für detailliertere Auswertungen müsste zunächst die Genauigkeit der Ergebnisse untersucht werden.

4. Erstellung der Statistik, Datenaufarbeitung, qualitätssichernde Maßnahmen

Datenerfassung

CAPI (*computer assisted personal interviewing*).

⁶ Hohe Besiedlungsdichte: eine Gruppe aneinander grenzender Gemeinden mit jeweils mehr als 500 Einwohner pro km² und insgesamt mindestens 50.000 Einwohner.

Mittlere Besiedlungsdichte: eine Gruppe aneinander grenzender Gemeinden mit 101-500 Einwohner pro km² und insgesamt mindestens 50.000 Einwohner.

Geringe Besiedlungsdichte: alle übrigen Gemeinden.

⁷ Dabei werden folgende Ausprägungen unterschieden: Wien, Gemeinden mit über 100.000 Einwohner, Gemeinden mit über 10.000 Einwohner, sonstige Gemeinden (d.s. alle Gemeinden mit maximal 10.000 Einwohner).

Signierung (Codierung)

Keine.

Plausibilitätsprüfung, Prüfung der verwendeten Datenquellen

Es zählt zu den großen Vorteilen einer CAPI-Erhebung, dass Erfassungsfehler (z.B. bei Betragsangaben eine 0 zu viel) sowie technisch unplausible Antwortmuster (z.B. wechselseitig nicht übereinstimmende Angaben zur Verwandtschaft) zum Teil bereits während der Befragung korrigiert werden können. Dies ist möglich durch die elektronische Datenerfassung und die integrierte Durchführung von Haushalts- und Personeninterviews. Mehr als 150 derartige Checks wurden von Statistik Austria entwickelt und von IFES implementiert.

Zu den [Plausibilitäts-Checks](#) zählen Warnings und Errors. Warnings sind beispielsweise Ober- und Untergrenzen beim Bezug bestimmter staatlicher Transfers. Gibt der Respondent einen Betrag an, der über einer Obergrenze bzw. unter einer Untergrenze liegt (oder vertippt sich der Interviewer), dann muss noch einmal nachgefragt werden. Ein Error ist eine Fehlermeldung, die der Interviewer auf dem Bildschirm bekommt, wenn eine Antwort aus dem Zusammenhang der Daten nicht möglich ist, etwa wenn Person A angibt, mit Person B verheiratet zu sein, Person B sich jedoch als geschieden bezeichnet.

Die MitarbeiterInnen der IFES-Zentrale führten ihrerseits Datenchecks durch, um die Umwandlung der Daten vom Erhebungsformat ins ASCII-Format und die Qualität der Interviews zu prüfen. Dazu erhielten sie regelmäßig Rückmeldung von Statistik Austria, wo die Daten dann eingehend plausibilisiert wurden.

In der Vorplaus wurde zunächst die Vollständigkeit der übermittelten Datenfiles kontrolliert. Anschließend wurden die verschiedenen Interviews in einem Haushalt auf Konsistenz geprüft und weitere technische Prüfungen vorgenommen.

Für die Mikroplaus wurden eigene SPSS-Programme geschrieben, wobei sich zeigte, dass diese inhaltlichen Datenprüfungen doch sehr aufwendig waren. Insbesondere die Bearbeitung der häufig auftretenden unplausiblen oder fehlenden Antworten auf Einkommensfragen erwies sich als sehr zeitaufwendig. Dazu ist anzumerken, dass nicht klar ist, wie viele dieser Antworten tatsächlich vom RespondentInnen kamen und in wie vielen Fällen es sich um InterviewerInnen-Fehler handelt.

Die Makroplaus erfolgte in erster Linie über Häufigkeitsauszählungen und Kohärenzprüfungen mit externen Datenquellen. Dabei erwies sich das Vorhandensein der Daten aus dem ECHP, aus der Konsumerhebung und aus den Steuerstatistiken als sehr hilfreich. So konnte, noch während der Feldarbeit, festgestellt werden, dass die Zahl der Personen mit einem Einkommen aus unselbständiger Erwerbstätigkeit zu niedrig lag, und dass sich diese Untererfassung regional auf bestimmte Bezirke konzentrierte. Tatsächlich lagen hier InterviewerInnen-Fehler vor, worauf in den betreffenden Haushalten eine telefonische Nacherhebung durchgeführt wurde.

Nach dem Vorliegen des authentischen internen Datenbestandes konnten die Zielvariablen gebildet werden. Um diese zu checken hatte Eurostat eigene SAS-Programme entwickelt, die auch den Mitgliedstaaten zur Verfügung gestellt worden waren.

Imputation (bei Antwortausfällen bzw. unvollständigen Datenbeständen)

Imputiert wurden fehlende Werte (*item non response*) bei Einkommensfragen. Diese können auf drei Arten entstehen: Die Zielperson verweigert die Angabe, ob sie eine bestimmte Einkommenskomponenten überhaupt bezogen hat (*Bezugsmissing*), es fehlt die Auskunft über die Zahl der Bezugsmonate oder es konnte oder wollte die Einkommenshöhe nicht genannt werden (*Betragsmissing*).

Die Codierung der Daten war so angelegt, dass Felder mit Item-Non-Response als Ausprägung immer „-1“ aufwiesen, während nicht zutreffende Felder (Leermeldungen) mit „-2“ codiert waren.

Bei den Bezugsmissings wurde die Ausprägung nach Möglichkeit deduktiv aus dem Zusammenhang der Daten imputiert, d.h. es wurden die verfügbaren Informationen über den Lebensunterhalt im Referenzjahr (aus dem „Aktivitätenkalender“) herangezogen, um den Bezug von Erwerbseinkommen, Arbeitslosengeld etc. festzulegen. Innerhalb einer Einkommensart (z.B. Pensionen) wurde maximal ein Bezug auf Ja gesetzt. Grundsätzlich auf Nein gesetzt wurden dagegen die Bezüge jener Einkommenskomponenten, bei denen eine deduktive Imputation nicht möglich war (z.B. Vermögenseinkünfte).

Fehlte die Monatsangabe, dann wurde sie nach Möglichkeit ebenfalls aus dem Zusammenhang der Daten rekonstruiert. Wo dies nicht der Fall war, wurde nach dem Muster der empirischen Verteilung eine Zufallszahl imputiert.

Betragsmissings sollten dadurch reduziert werden, dass im Falle von Unwissenheit des genauen Betrages die RespondentInnen eine Karte vorgelegt bekamen, auf denen Einkommensstufen (Intervalle) angegeben waren. Diese Stufenangaben wurden dann ebenfalls durch eine in ihrer Verteilung der empirischen Verteilung entsprechende Zufallsvariable ersetzt.

Fehlte bei einer Einkommenskomponente, die sowohl brutto als auch netto anzugeben war, ausschließlich einer der beiden Einkommenswerte, so wurde der korrespondierende Brutto- bzw. (seltener) Nettobetrag in einem Modell berechnet, das sich auf die österreichischen Steuerdaten stützt (siehe w. u.).

Hatten RespondentInnen nicht einen konkreten Betrag, sondern ein Betragsintervall (Stufe) angegeben, so erfolgte eine stochastische Imputation auf Basis der beantworteten Fälle. D.h. es wurden Zufallszahlen imputiert, die in ihren Verteilungen den relativen Häufigkeiten der gültigen Fälle im jeweiligen Intervall entsprachen.

Die Imputation der Betragsmissings erfolgte in einem Mittelweg aus deterministischer und stochastischer Imputation. Als Schätzverfahren wurden komponentenweise lineare Regressionsmodelle gewählt, wobei das Einkommen zwecks besserer Annäherung an die Normalverteilung logarithmiert wurde. Um die Varianz nicht zu unterschätzen wurden die Prognosewerte aus den Regressionen um fiktive Residuen ergänzt, d.h. die Addition dieser stochastischen Störterme (sie entsprechen bei der Varianzzerlegung der Variabilität innerhalb der Gruppen) sollte eine valide Abbildung der Verteilung des interessierenden Merkmals gewährleisten.

Beispiel: Logarithmiertes Nettoeinkommen aus unselbständiger Erwerbstätigkeit im Jahr 2002.

Fehlende Ausprägungen wurden mit folgenden Prädiktoren imputiert

- Geschlecht (Faktor)
- Alter (Kovariate)
- höchste abgeschlossene Ausbildung (F)
- logarithmierte Zahl der bisherigen Berufsjahre (K)
- Qualifikationsebene der ausgeübten Tätigkeit (F)
- Zahl der normalerweise pro Woche geleisteten Arbeitsstunden (K)
- Logarithmus des laufenden monatlichen Bruttoeinkommens (K)
- Bundesland (F, zusammengefasst)
- Branche (F, zusammengefasst)
- berufliche Tätigkeit (F, zusammengefasst)
- Zahl der Beschäftigungsmonate im Jahr 2002 (K)
- Lehrling ja/nein (dichotom)
- Wechselwirkung zwischen Geschlecht, sozialer Stellung (F) und Alter

Insgesamt wurden bei dieser Variable 847 fehlende Werte imputiert, darunter 246 mit sämtlichen Prädiktoren und 458 mit sämtlichen Prädiktoren außer dem Logarithmus des laufenden monatlichen Bruttoeinkommens (da dieses ebenfalls unbekannt war).

In die Regressionen flossen gleichsam metrische wie kategorielle Variablen ein. Die Auswahl der Regressoren erfolgte einerseits nach deren Erklärungsgehalt (Zuwachs beim korrigierten R^2), andererseits nach deren inhaltlicher Interpretierbarkeit. Kategorielle Variablen wurden nach der Dummy-Methode codiert. Nachdem die Regressoren teilweise selbst Item-Non-Response aufwiesen, mussten manchmal mehrere Modelle gerechnet werden.

Gegenüber einer stochastischen Imputation (z.B. Hot-decking) hat die gewählte Form des linearen Regressionsmodells den Nachteil einer zeitlich aufwendigen Modellspezifikation für jede betroffene Einkommenskomponente. Dem steht aber der beträchtliche Vorteil gegenüber, dass ein lineares Regressionsmodell bereits bei verhältnismäßig kleinen Fallzahlen anwendbar ist und die Item-Non-Response-Quote nicht so stark zum Tragen kommt. Gerade in einer so detaillierten Komponentenerhebung wie bei EU-SILC, wo manche Komponenten nur von 30 oder 50 RespondentInnen bezogen werden, ist dieser Umstand wesentlich.

Allerdings konnte nicht immer ein lineares Regressionsmodell festgelegt werden. Auch erschien ein Regressionsmodell nicht immer sinnvoll, etwa beim Kindergeld, wo die Imputation des gesetzlich festgelegten Monatssatzes zu mutmaßlich besseren Ergebnissen führt. Bei den übrigen Einkommenskomponenten, auf die ein lineares Regressionsmodell nicht zufrieden stellend anwendbar war, wurde entweder eine Zufallszahl oder der Median ergänzt um eine Zufallszahl imputiert.

Hochrechnung (Gewichtung)

In Analogie zu der ab 2004 verbindlichen Weise wurden für Personen und Haushalte in EU-SILC-2003 jeweils drei Hochrechnungsgewichte ermittelt:

Das *Designgewicht* stellt die Umkehrung der Auswahlwahrscheinlichkeit dar. Dabei war im zweistufigen Bereich die Klumpung zu berücksichtigen.

Das *Non-Response-Gewicht* kompensiert den Unit-Non-Response auf Haushalts- wie auf Personenebene und entspricht dem Quotienten aus Bruttostichprobe durch Nettostichprobe. Als Gliederungskriterien wurde die Verkreuzung von Bundesland und Besiedlungsdichte (24 Ausprägungen) verwendet. Dadurch sollte sichergestellt werden, dass regional unterschiedliche Antwortquoten in Kombination mit regional unterschiedlichen Einkommensniveaus nicht die Schätzung der Einkommensverteilung im Bundesgebiet verzerren.

Das *Anpassungsgewicht* zur Anpassung von Randverteilungen an die Populationswerte wurde in einem iterativen Prozess generiert (*Raking*). Verwendete Merkmale auf Personenebene waren Geschlecht, Bundesland und Altersgruppen (6 Ausprägungen⁸), auf Haushaltsebene wurde an das Bundesland und die Haushaltgröße (5 Ausprägungen) angepasst. Diese Vorgangsweise entspricht im Wesentlichen den Empfehlungen von Eurostat für die Erhebungen ab 2004, die gewählten Merkmale bzw. Ausprägungen stellen einen Kompromiss dar zwischen der Übereinstimmung mit den Populationswerten und der Variation in den Hochrechnungsgewichten.

Als Datenquellen für die beim Raking verwendeten Personenmerkmale dienten die Bevölkerungsfortschreibung vom Dezember 2002 (Gesamtbevölkerung) sowie der Mikrozensus Dezember 2002 (Bevölkerung in Anstaltshaushalten; die Bevölkerung in Privathaushalten ist somit eine Residualgröße). Ebenfalls aus dem Mikrozensus Dezember 2002 stammt die Verteilung der Haushaltgröße von Privathaushalten.

Erstellung des Datenkörpers, (weitere) verwendete Rechenmodelle, statistische Schätzmethoden

Die Berechnung von Familienbeihilfe/Kinderabsetzbetrag erfolgte entsprechend den gesetzlich festgelegten Sätzen.

Ein weiteres Rechenmodell betraf die Zuordnung von Einmalzahlungen wie Abfertigungen oder Urlaubsentschädigungen. Die Antwort auf die Frage, inwieweit solche Zahlungen Teil des laufenden Einkommens sind, ist international umstritten. Statistik Austria rief daher Eurostat an, eine verbindliche Regelung auszugeben. Nachdem von dort lange Zeit keine Antwort kam, wurden die Einmalzahlungen in einem nationalen Verfahren zugeordnet. Sie wurden anteilig

⁸ Verwendet wurden: 0-15 Jahre, 16-29 Jahre, 30-39 Jahre, 40-49 Jahre, 50-64 Jahre, 65 und mehr Jahre.

entweder als Unselbständigen-, Arbeitslosen- oder Pensionseinkommen klassifiziert, je nach aktuellem Erwerbsstatus der Zielperson.

Neben den Einmalzahlungen waren zwei der österreichischen Erhebungsvariablen den Eurostat-Zielvariablen nicht eindeutig zuordenbar. Das Einkommen aus Präsenz- und Zivildienst wurde als Einkommen aus unselbständiger Erwerbstätigkeit klassifiziert. Das „sonstige Einkommen“ wurde je nach Plausibilität dem Unselbständigen-, Selbständigen- oder Pensionseinkommen zugerechnet.

Bei PensionistInnen war nach dem Monatseinkommen gefragt worden. Um die Nettowerte auf Jahresbasis bringen zu können, musste die Steuerbegünstigung von Sonderzahlungen rechnerisch inkludiert werden. Dies geschah unter Entwicklung eines entsprechenden Faktors auf Basis der Lohnsteuerdaten 2002.

Ein weiteres angewandtes Rechenmodell war die Brutto-Netto-Rechnung bzw. Netto-Brutto-Rechnung. Diese wurde notwendig, wenn eine Zielperson entweder nur das Brutto- oder – was häufiger der Fall war – nur das Nettoeinkommen angegeben hatte. Die Umrechnung erfolgte bei Unselbständigen und PensionistInnen auf Basis der Lohnsteuerdaten 2002, also auf Basis realer Faktoren. Für selbständig Erwerbstätige sind solche Informationen nicht vorhanden; hier mussten auf Basis der Einkommensteuerdaten 2000 Faktoren für die Einkommensteuer geschätzt und die Sozialversicherungsbeiträge in einem Modell berechnet werden.

Zur Brutto-Netto-Rechnung bzw. Netto-Brutto-Rechnung ist generell zu sagen, dass es sich hierbei um einen sehr komplexen Rechenvorgang handelt, da ein und dieselbe Person mehrere Einkommenskomponenten beziehen und dabei auch unterschiedliche Antwortmuster aufweisen kann.

Informationen zur Imputation von Antwortausfällen sind unter dem Punkt „Imputation bei Antwortausfällen“ w. o. zu finden.

Sonstige qualitätssichernde Maßnahmen

Im Mai/Juni 2002 hatte eine Piloterhebung in 200 Haushalten stattgefunden. Die Feldarbeit war von Fessel-GfK durchgeführt worden, wobei einerseits die Funktionalität der CAPI-Erhebung getestet worden war. Andererseits lieferten die Ergebnisse der Piloterhebung auch wertvolle Anregungen zur Optimierung des Fragebogens und der vorgegebenen Plausibilitäts-Checks. Durch ein InterviewerInnen-Feedback sowie ein Debriefing mit den zuständigen InstitutsmitarbeiterInnen konnten diesbezüglich weitere Erkenntnisse gewonnen werden.⁹

Dementsprechend wurden zahlreiche Erläuterungen zum Fragebogen bereitgestellt, die in das von IFES letztlich verwendete Manual größtenteils Eingang nahmen. Auch nahmen MitarbeiterInnen der Statistik Austria an den obligatorischen InterviewerInnenschulungen teil, insbesondere in der Anfangsphase. (Dadurch konnten auch die Schulungsunterlagen für die Folgetermine noch kurzfristig verbessert werden.)

Um den Einfluss einer selektiven Adressenbearbeitung auf die Qualität der Ergebnisse einzuschränken wurden in die Ausschreibung folgende Bestimmungen aufgenommen: Es sollten pro Haushalt zumindest 3 Kontaktversuche erfolgen, bevor eine Verweigerung akzeptiert werden konnte, und es mussten – neben den bundesweit 60% – in jedem Bezirk zumindest 50% der bearbeiteten Haushalte erfolgreich befragt werden. Insbesondere letztere Bestimmung führte zu einem hohen Einsatz des Instituts, auch als „schwierig“ empfundene Haushalte (v.a. im städtischen Bereich) zur Mitarbeit zu bewegen, was nicht zuletzt durch den Einsatz von Incentives (v.a. Gutscheine) gelang.

Als höchst wirksames Instrument zur diesbezüglichen Qualitätssicherung erwiesen sich die Feldberichte. In der Ausschreibung war festgelegt worden, dass während der Feldzeit alle 2 Wochen eine Datei an Statistik Austria zu übermitteln war, die eine Liste aller bis dorthin er-

⁹ Vgl. Statistik Austria: „EU-SILC-Piloterhebung Österreich. Endbericht einer im Auftrag von Eurostat durchgeführten Erhebung“, Wien 2002.

folgten Kontaktversuche mit dem Kontaktergebnis beinhaltete. Auf diese Weise konnten allfällige Fehlentwicklungen in der Feldarbeit bereits frühzeitig erkannt und korrigiert werden.

Als inhaltliche Qualitätssicherungsmaßnahme seien schließlich die Zwischendatenlieferungen angeführt. IFES erhielt jedes Mal umfassende Rückmeldung zu den Daten, was zur Information der InterviewerInnen und zur Nachbefragung in einzelnen Haushalten führte, wodurch die Qualität der endgültigen Daten weiter gesteigert werden konnte.

5. Publikation (Zugänglichkeit)

Vorläufige Ergebnisse

Keine.

Endgültige Ergebnisse

Übermittlung der Mikrodaten an Eurostat am 30.7.2004 (zusammen mit dem englischsprachigen, standardisierten EU-Qualitätsbericht), weiters Übermittlung der auf Basis dieser Indikatoren berechneten Laeken-Indikatoren am 29.10.2004 (zusammen mit dem Abschlussbericht).

Auf nationaler Ebene wurden Hauptergebnisse zu Einkommensverteilung und Armut tabellarisch (im Statistischen Jahrbuch) ab Dezember 2004, kommentiert (d.h. als Publikation) ab Februar 2005 veröffentlicht.

Revisionen

Trifft nicht zu.

Publiziert in:

Statistischer Annex zum 2005 Bericht der Kommission für die Tagung des Europäischen Rates

[Statistisches Jahrbuch Österreichs](#)

Kapitel 9, Tabellen 9.27 und 9.28.

[Statistische Nachrichten](#)

- *EU-SILC - die neue Erhebung zu Einkommen und Lebensbedingungen*, Heft 3/2005, ISSN 0029-9960, Art.Nr. 20-0803-05, S. 224 ff.
- *Messung von Armutgefährdung und Deprivation*, Heft 4/2005, ISSN 0029-9960, Art.Nr. 20-0804-05, S. 349 ff.

Standard-Publikation

[Einkommen, Armut und Lebensbedingungen. Ergebnisse aus EU-SILC 2003 in Österreich.](#)

Statistik Austria (inklusive CD-ROM), ISBN 3-902479-08-6, Art.Nr. 20-1860-03, Wien 2005.

Armut und Armutgefährdung in Österreich 2003,

erschieden im [Bericht über die soziale Lage in Österreich 2003 - 2004](#),

hrsg. v. BMSG, ISBN 3-85010-132-0, Wien 2005, S. 207 ff.

Internet

[Homepage der Statistik Austria - EU-SILC](#)

Behandlung vertraulicher Daten

Um die vom Erhebungsinstitut gelieferten Daten auf Vollständigkeit und Richtigkeit prüfen zu können, waren Familien- und Vorname der Person sowie das vollständige Geburtsdatum zu erheben. Weiters lag aus der Stichprobe die Adresse vor. Diese Merkmale waren Teil des Registers in der Erhebungsphase, also auf einem anderen Datensatz gespeichert als die

inhaltlichen Erhebungsdaten wie Einkommen, Gesundheit usw. Nach Abschluss der Plausibilitätsprüfungen wurden Name, Adresse und Geburtstag gelöscht. Im Datensatz verblieben lediglich Geburtsmonat, Geburtsjahr und das Bundesland (NUTS2), da es sich hierbei um Eurostat-Zielvariablen handelt.

Am Eurostat-Datenfile selbst sind die Haushalte durch eine Laufnummer (5-stellig) gekennzeichnet. Die Identifikation einer Person erfolgt über eine 7-stellige Personennummer, wobei es sich bei den ersten 5 Stellen um die Laufnummer des Haushalts, bei den letzten 2 Stellen um eine Personen-Laufnummer innerhalb des Haushalts handelt.

Die Geheimhaltungsbestimmungen für Daten, die im Bundesstatistikgesetz 2003 konsolidierte Fassung §19 (2) und (3) geregelt sind, werden strikt eingehalten.

6. Qualität

6.1. Relevanz

EU-SILC 2003 wurde auf Grundlage von bilateralen Verträgen mit Eurostat und dem österreichischen Sozialministerium durchgeführt. Es fand eine intensive Kommunikation mit den Auftraggebern statt. Durch die fristgerechte und einwandfreie Erfüllung dieser Verträge sind Interessen der beiden Hauptnutzer erfüllt. Seitens des Sozialministeriums gab es einen Zusatzauftrag zur Erstellung eines Kapitels im Sozialbericht zu Armut und Armutsgefährdung in Österreich auf Grundlage der EU-SILC 2003 – Daten.

Es wird häufig Interesse an regionalen Auswertungen bekundet. Da eine – fachlich wünschenswerte – Vergrößerung der Stichprobe für detailliertere Auswertungen in regionaler und inhaltlicher Hinsicht, insbesondere auch im Längsschnitt, aus EU- oder Bundesmitteln bedauerlicherweise nicht finanzierbar war, wurden im Rahmen der Ausschreibung der Feldarbeit für die Jahre 2003 und 2004 auch Kostenvoranschläge für eine allfällige Aufstockung der Stichprobe eingeholt und diese den Ländern und anderen Interessierten zur Kenntnis gebracht.

6.2. Genauigkeit

6.2.1. Stichprobenbedingte Effekte, Repräsentativität

Das arithmetische Mittel des äquivalisierten Netto-Haushaltseinkommens liegt bei EUR 17.620. Der Schätzwert für den Standardfehler dieses Parameters liegt bei EUR 180 oder 1,0%. Das bedeutet, dass (unter Vernachlässigung systematischer Fehler) der tatsächliche Erwartungswert dieser Größe mit einer Wahrscheinlichkeit von 95% im Intervall von EUR 17.270 bis EUR 17.970 liegt.

Särndal et. al (1992, S. 197 ff.) haben ein nichtparametrisches Verfahren zur Kalkulation der Varianz von gewichteten Quantilen (und damit auch des gewichteten Medians) entwickelt. Dabei wird von Konfidenzbändern der kumulierten Verteilungsfunktion (CDF) ausgegangen. Der Wert der CDF an einer bestimmten Stelle x^* wiederum wird betrachtet als das arithmetische Mittel von n Indikatorvariablen (eine pro Beobachtung), die das standardisierte Hochrechnungsgewicht annehmen, falls $x_i \leq x^*$ und 0 sonst ($i = 1, \dots, n$). Die x -Werte der Schnittpunkte der solcherart bestimmten Konfidenzbänder der CDF mit der 50%-Horizontalen entsprechen den Grenzen des Vertrauensbereichs für den gewichteten Median.

Im Gegensatz zum arithmetischen Mittel existiert für den Median (die wichtigste eindimensionale Maßzahl zur Darstellung von Einkommensverteilung) keine unmittelbar anwendbare arithmetische Formel zur Berechnung von Konfidenzintervallen für den aus der Stichprobe geschätzten Populationsparameter. Dies gilt insbesondere bei der Verwendung von Hochrechnungsgewichten, da in diesem Fall auch die ordinale Abschätzung eines Vertrauensbereichs über die Binomialverteilung unzulässig ist.

Mit diesem Verfahren ergibt sich für den Median des äquivalisierten Netto-Haushaltseinkommens bei einem Schätzwert von EUR 15.710 ein Standardfehler von EUR 168 oder 1,1%. In

einer weiteren Analyse könnte noch ventiliert werden, wie gut die zur Anwendung dieser Methode erforderlichen Verfahrensvoraussetzungen erfüllt sind bzw. ob bei einer Anwendung von Replikationsmethoden (etwa Bootstrap-t) andere Ergebnisse zu erwarten wären.

Der Designeffekt ist eine Maßzahl für die Variabilität von Parametern im Vergleich zur entsprechenden Variabilität bei einer einfachen Zufallsstichprobe ohne Antwortausfälle. In EU-SILC 2003 wird der Designeffekt beeinflusst durch die geklumpete Stichprobe, die Klumpung aller Personen innerhalb eines Haushalts und die gebundene Hochrechnung (diese Entzerrung ist notwendig, weil die Nettostichprobe auf Grund selektiven Ausfalls keine hinreichende Repräsentativität hinsichtlich demographischer Basismerkmale gewährleisten würde). In Bezug auf das mittlere äquivalisierte Netto-Haushaltseinkommen beträgt der Designeffekt 1,71, d.h. der Standardfehler dieser Größe ist 71% größer als in der fiktiven Referenzgruppe.

Aus der realisierten Stichprobengröße von 11.991 Personen und dem Designeffekt von 1,71 ergibt sich eine effektive Stichprobengröße von 7.012 Personen in Bezug auf das mittlere äquivalisierte Netto-Haushaltseinkommen.

6.2.2. Nicht-stichprobenbedingte Effekte

Qualität der verwendeten Datenquellen

Nicht zutreffend, da eigene Datenerhebung.

Abdeckung (Fehlklassifikationen, Unter-/Übererfassung)

Die Stichprobe entstammt dem Zentralen Melderegister (ZMR). Dieses war nach der letzten Volkszählung durch die Zusammenführung der Gemeinde-Melderegister erstmals initialisiert worden und wird seither, auf Basis von Wohnsitzmeldungen, laufend aktualisiert. Die Administration obliegt dabei dem Innenministerium, das ZMR umfasst Haupt- und Nebenwohnsitze.

Leider war bei der Erstbefüllung des ZMR nicht auf den Eintrag eines eindeutigen Wohnungsidentifikators geachtet worden, sodass ein Wohnungszusammenhang nur über die Adresse herstellbar ist. Auch variierte die Qualität der Einträge in den einzelnen Gemeinden erheblich, weshalb in der ZMR-Aufbauphase umfassende Cleanings und Harmonisierungen notwendig wurden, die zum Zeitpunkt der Stichprobenziehung für EU-SILC 2003 erst teilweise umgesetzt worden waren.

Die Herstellung eines Wohnungszusammenhangs (und damit eines Haushaltszusammenhangs) über die Adresse funktioniert nicht immer einwandfrei (fehlende Türnummern, unterschiedliche Schreibweisen usw.). Als ein Abdeckungsfehler ist daher das Nicht-Erkennen von Wohnungszusammenhängen und ein daraus resultierender verzerrter Auswahlrahmen zu nennen. Das wird dadurch belegt, dass die Zahl der Einpersonenhaushalte im ZMR höher ausfällt als z.B. im Mikrozensus. Weniger wahrscheinlich ist ein irrtümliches Zusammenfassen von Personen, die nicht in einer Wohnung leben.

Weiters können, bedingt durch die Konstruktion des ZMR auf Basis des Meldewesens, klassifikatorische Fehler auftreten, wenn die Hauptwohnsitzmeldung einer Person nicht dem tatsächlichen Mittelpunkt der Lebensbeziehungen entspricht und reale Hauptwohnsitze im ZMR nur als Nebenwohnsitze oder gar nicht aufscheinen.

Daneben muss der Zeitabstand zwischen der Stichprobenziehung und der Feldarbeit berücksichtigt werden: Nachdem das ZMR kontinuierlich aktualisiert wird, entspricht der Auswahlrahmen zum Zeitpunkt der Stichprobenziehung zumindest formal zwar tatsächlich der korrespondierenden Population. Da aber bis zur konkreten Datenerhebung mitunter einige Monate vergehen, ist auch hier mit geringfügigen Verzerrungen zu rechnen (Gebäudeabrisse, Neubauten, Umzüge usw.).

Gegenüber anderen Fehlern haben Abdeckungsfehler den Vorteil, dass sie während der Feldphase zumindest teilweise korrigiert werden können. So wurden etwa nicht (mehr) existente Adressen als qualitätsneutrale Ausfälle verbucht. Auch konnte der Haushaltszusammenhang der einzelnen Mitglieder eines Haushalts durch die InterviewerInnen in der Regel korrekt erfasst werden.

Antwortausfall (Unit-Non-Response, Item-Non-Response)

Unit-Non-Response

Von den 8931 ausgegebenen Adressen wurden 7340 vom Erhebungsinstitut tatsächlich verwendet, d.h. diese Bruttostichprobe reichte aus, um netto 4500 Haushalte zu erzielen.

7106 Adressen konnten kontaktiert werden; der Rest entfiel auf nicht existente Adressen (n = 158) sowie auf unauffindbare oder nicht zugängliche Adressen (n = 76). Die Adressenkontaktierungsquote beträgt somit 98,9%.

Aus den 7106 Adressen wurden vom Erhebungsinstitut 4725 Haushalte ausgeschöpft, 2381-mal (33,5%) kam es zu Unit-Non-Response. Darunter entfiel mehr als die Hälfte auf Verweigerungen und ein Drittel auf „Niemand angetroffen“.

Von den 4725 gelieferten Mikrodatsätzen wurden von Statistik Austria 102 (2,2%) im Zuge der Plausibilitätskontrollen als unbrauchbar ausgeschieden. Eine solche Unbrauchbarkeit lag insbesondere dann vor, wenn

- keinerlei Information vorhanden war, ob Einkommen bezogen worden war oder nicht,
- mehr als 50% der erforderlichen Personeninterviews fehlten oder
- der gewichtete Item-Non-Response in Summe einen bestimmten Grenzwert überschritt.

Es flossen somit 4623 Haushalte in die Datenauswertung und -analyse ein. Die Antwortquote auf Haushaltsebene beträgt 64,4% (= 4623/7182) bzw. der Unit-Non-Response als Residualgröße 35,6%.¹⁰

Interview-Ausschöpfung auf Haushaltsebene

	absolut	in %
Ausgegebene Adressen	8931	100.0
Vom Erhebungsinstitut verwendet	7340	82.2
Vom Erhebungsinstitut nicht verwendet	1591	17.8
Verwendete Adressen	7340	100.0
Adresse existent	7182	97.8
Adresse nicht existent	158	2.2
Bruttostichprobe	7182	100.0
Adresse erfolgreich kontaktiert	7106	98.9
Adresse nicht erfolgreich kontaktiert	76	1.1
Erfolgreich kontaktierte Adressen	7106	100.0
Haushaltsinterview erfolgreich durchgeführt	4725	66.5
Verweigerung	1267	17.8
Niemand anwesend	799	11.2
Mitarbeit wegen Krankheit etc. nicht möglich	60	0.8
Sonstiger Ausfallsgrund	255	3.6
Erfolgreich durchgeführte Haushaltsinterviews	4725	100.0
Nach Plausibilitätskontrollen akzeptiert	4623	97.8
Nach Plausibilitätskontrollen nachträglich ausgeschlossen	102	2.2

¹⁰ Entsprechend den Vorgaben von Eurostat werden nicht existente Adressen als nicht zur Grundgesamtheit gehörig und somit für die Ausschöpfungsquote irrelevant betrachtet, während unauffindbare und nicht zugängliche Adressen wie Verweigerungen behandelt werden.

Wie unten stehende Tabelle zeigt, war diese Unit-Non-Response-Quote regional unterschiedlich, allerdings in vertretbarem Ausmaß. Dazu hat die Ausschreibungsvorgabe der Mindestantwortquote von 50% pro Bezirk sicher wesentlich beigetragen.

Entwicklung der Unit-Non-Response-Quote auf Haushaltsebene

Feldarbeitszeit bzw. Projektstand	Burgenland	NÖ	Wien	Kärnten	Steiermark	OÖ	Salzburg	Tirol	Vorarlberg	TOTAL
04.06. - 17.06.	36.4	26.1	71.2	0.0	18.6	44.6	0.0	65.1	34.2	40.9
04.06. - 01.07.	38.5	55.0	77.4	40.5	29.8	43.7	56.1	47.1	37.9	51.4
04.06. - 15.07.	32.1	56.1	73.4	41.2	32.6	46.6	43.4	47.6	36.0	51.0
04.06. - 29.07.	27.1	52.6	70.4	33.7	33.8	44.3	40.0	47.3	44.0	48.3
04.06. - 12.08.	27.8	48.1	58.9	36.2	33.5	42.1	35.4	40.8	44.2	44.3
04.06. - 26.08.	27.2	43.8	53.4	32.0	31.1	39.4	34.9	36.6	50.6	40.9
04.06. - 09.09.	24.7	41.5	49.0	31.1	29.7	36.3	30.2	32.8	48.4	38.1
04.06. - 23.09.	22.5	40.3	44.6	31.3	26.3	31.7	27.9	29.1	48.7	35.3
04.06. - 07.10.	26.6	37.9	43.5	32.0	27.8	30.9	30.8	30.1	38.1	34.8
04.06. - 17.10.	26.6	37.5	41.7	31.9	27.8	31.0	30.8	30.1	38.1	34.3
Nach telefon. Nachbefragung	26.6	37.5	41.6	31.9	27.8	30.4	30.8	30.1	38.1	34.2
Nach Plausibilitätskontrollen	27.6	38.8	42.6	32.7	29.0	31.9	32.8	33.0	40.5	35.6

Es fällt auf, dass in der ersten Hälfte der Feldarbeit (abgesehen vom Sonderfall der ersten 2 Wochen) der Anteil des Unit-Non-Response in allen Bundesländern kontinuierlich zurückgeht. Das ist primär eine Folge der mehrfachen Kontaktierung von Haushalten, die nicht angetroffen worden waren (durch dieselben InterviewerInnen) bzw. die eine Mitarbeit verweigert hatten (durch andere InterviewerInnen). Ab der zweiten Hälfte der Feldarbeit konzentrierte sich das Erhebungsinstitut im Wesentlichen auf die „Problembezirke“, das waren jene mit bis dahin zu geringer Ausschöpfung. Dies erklärt die erfolgreiche Reduktion der Quoten in diesem Zeitraum vor allem in Wien und Vorarlberg. Interview-Abbrüche auf Grund erfüllter Soll-Quoten (durch IFES) sowie die Elimination unbrauchbarer Haushalte (durch Statistik Austria) führten dazu, dass am Ende der Unit-Non-Response-Anteil in den Ländern zumeist wieder etwas anstieg. Die folgende Tabelle zeigt die Adressausschöpfung nach Bundesländern in absoluten Zahlen.

	Burgenland	NÖ	Wien	Kärnten	Steiermark	OÖ	Salzburg	Tirol	Vorarlberg	Österreich
Ausgegebene Adressen	285	1665	2035	582	1270	1482	558	695	359	8931
Nicht verwendete Adressen	82	260	375	93	210	216	145	146	64	1591
Verwendete Adressen	203	1405	1660	489	1060	1266	413	549	295	7340
Qualitätsneutrale Ausfälle	11	30	46	3	28	23	1	10	6	158
Gültige Adressen	192	1375	1614	486	1032	1243	412	539	289	7182
Nicht erreichte Adressen	1	8	24	7	10	11	1	5	9	76
Erreichte Adressen	191	1367	1590	479	1022	1232	411	534	280	7106
Niemand anwesend	20	189	250	37	85	100	41	54	23	799
Mitarbeit verweigert	26	209	282	94	154	206	79	78	56	1184
Sonstiger Ausfallsgrund	4	109	116	17	38	61	6	25	22	398
Durchgeführte Befragungen	141	860	942	331	745	865	285	377	179	4725
Zurückgewiesene Befragungen	2	19	15	4	12	19	8	16	7	102
Akzeptierte Befragungen	139	841	927	327	733	846	277	361	172	4623
Ausschöpfungsquote (%)	72.4	61.2	57.4	67.3	71.0	68.1	67.2	67.0	59.5	64.4

Personenebene: In den 4623 schlussendlich verwendeten Haushalten lebten 11991 Personen, davon 9624 mit Geburtsjahr 1986 oder früher (d.h. waren zu befragen). In 9543 Fällen (99,2%) kam ein Personeninterview zu Stande; davon entfielen 14,4% auf Proxy-Interviews.

Interview-Ausschöpfung auf Personenebene

	absolut	in %
Zahl der Personen in der Stichprobe	11991	100.0
davon 1986 oder früher geboren (d.h. zu interviewen)	9624	80.3
davon 1987 oder später geboren (d.h. nicht zu interviewen)	2367	19.7
Zu interviewende Personen	9624	100.0
Interview erfolgreich durchgeführt	9543	99.2
Krankheit etc. und Proxy-Interview nicht möglich	7	0.1
Verweigerung	35	0.4
Abwesenheit und Proxy-Interview nicht möglich	36	0.4
Sonstiger Ausfallsgrund	3	0.0
Befragungstechnik der erfolgreichen Interviews	9543	100.0
Persönliche Befragung (Selbstauskunft)	8170	85.6
Proxy-Interview (Fremdauskunft)	1373	14.4

Der Unit-Non-Response auf Personenebene fällt mit 0,8% sehr gering aus. Das hat zwei Ursachen: Erstens waren eben in bestimmten Fällen Proxy-Interviews gestattet, und zweitens wurden solche Haushalte, bei denen Personeninterviews fehlten, weit eher nachträglich aus dem Datensatz ausgeschlossen.

Item-Non-Response

Item-Non-Response ist in erster Linie bei Einkommensfragen von Relevanz. Nachdem das Haushaltseinkommen aus verschiedenen Komponenten zusammengesetzt ist, ist es unbekannt, sobald bei einer Komponente Item-Non-Response vorliegt.

Unten stehende Tabelle gibt an, wie viele Haushalte bei den Einkommens-Zielvariablen nach Abschluss dieser Prozeduren noch Item-Non-Response aufwiesen. Nachdem die Zielvariablen teilweise aus mehreren Erhebungsvariablen bestehen, kann dies auch partiell auftreten.

Item-Non-Response bei den Zielvariablen

Zielvariable	Einkommensbezieher		Informationen über die Einkommenshöhe ...						
	absolut	in %	vollständig		teilweise vorhanden		nicht vorhanden		
			absolut	in %	absolut	in %	absolut	in %	
Gesamtes Haushaltseinkommen									
HY010	Brutto-Haushaltseinkommen	4621	100.0	2731	59.1	1618	35.0	272	5.9
HY020	Netto-Haushaltseinkommen	4621	100.0	2731	59.1	1618	35.0	272	5.9
HY022	Netto-Haushaltseinkommen ohne soziale Transfers (ausg. Alter und Hinterbliebene)	4504	97.4	2684	59.6	1313	29.2	507	11.3
HY023	Netto-Haushaltseinkommen ohne soziale Transfers	3657	79.1	2019	55.2	1017	27.8	621	17.0
Haushaltseinkommen ohne Vermögenseinkommen									
HY010	Brutto-Haushaltseinkommen	4621	100.0	3226	69.8	1117	24.2	278	6.0
HY020	Netto-Haushaltseinkommen	4621	100.0	3226	69.8	1117	24.2	278	6.0
HY022	Netto-Haushaltseinkommen ohne soziale Transfers (ausg. Alter und Hinterbliebene)	4490	97.1	3188	71.0	783	17.4	519	11.6
HY023	Netto-Haushaltseinkommen ohne soziale Transfers	3339	72.2	2258	67.6	622	18.6	459	13.7
Einkommenskomponenten auf Haushaltsebene									
HY040	Vermietung und Verpachtung	314	6.8	246	78.3	4	1.3	64	20.4
HY050	Familienleistungen	1762	38.1	1724	97.8	36	2.0	2	0.1
HY060	Sozialhilfe	101	2.2	94	93.1	1	1.0	6	5.9
HY070	Wohnbeihilfe/Mietzinsbeihilfe	111	2.4	100	90.1	0	0.0	11	9.9
HY080	Regelmäßig erhaltene Inter-Haushalt-Transfers (z.B. Unterhaltszahlungen)	281	6.1	239	85.1	1	0.4	41	14.6
HY090	Interests, dividends, profit from capital investments in unincorporated business	1221	26.4	472	38.7	84	6.9	665	54.5
HY110	Einkommen von Unter-16-Jährigen	36	0.8	25	69.4	0	0.0	11	30.6
HY130	Regelmäßig geleistete Inter-Haushalt-Transfers (z.B. Unterhaltszahlungen)	299	6.5	269	90.0	4	1.3	26	8.7
Einkommenskomponenten auf Personenebene									
PY010	Unselbständige Erwerbstätigkeit	4752	49.8	3724	78.4	149	3.1	879	18.5
PY035	Beiträge zu privater Pensionsvorsorge	1814	19.0	1495	82.4	0	0.0	319	17.6
PY050	Selbständige Erwerbstätigkeit	990	10.4	567	57.3	9	0.9	414	41.8
PY080	Privatpensionen	65	0.7	37	56.9	0	0.0	28	43.1
PY090	Arbeitslosenleistungen	576	6.0	486	84.4	2	0.3	88	15.3
PY100	Altersleistungen	2069	21.7	1758	85.0	9	0.4	302	14.6
PY110	Hinterbliebenenleistungen	599	6.3	515	86.0	0	0.0	84	14.0
PY120	Krankheitsleistungen	233	2.4	228	97.9	1	0.4	4	1.7
PY130	Invalideits-/Unfallsleistungen	530	5.6	459	86.6	5	0.9	66	12.5
PY140	Bildungsleistungen	132	1.4	106	80.3	2	1.5	24	18.2

Lesebeispiel: Von allen Haushalten bezogen 314, das sind 6.8%, ein Einkommen aus Vermietung und Verpachtung. Dabei war die Information über die Einkommenshöhe 246-mal (78.3% dieser 314) vollständig vorhanden, 4-mal teilweise vorhanden (z.B. bei einem Haushaltsmitglied, bei einem anderen nicht), und 64-mal nicht vorhanden. Die Prozentsätze bei vollständig/teilweise/nicht vorhanden summieren sich immer auf 100 Prozent, die entsprechenden Absolutzahlen auf den Wert der ersten Spalte. Bei Interpretation der Prozentsätze in Spalte 2 sollte bedacht werden, dass für Haushalte und Personen die Grundgesamtheiten unterschiedlich groß sind (4623 Haushalte bzw. 9543 Personen).

Besonders hoch ist der Item-Non-Response erwartungsgemäß bei Vermögenseinkünften, ferner beim Selbständigeneinkommen, bei den Privatpensionen und den Einkommen der Unter-16-Jährigen (wobei die letzten beiden quantitativ eher unbedeutend sind). Dagegen ist bei Sozialtransfers der Anteil an Antwortausfällen wesentlich geringer. Die niedrige Quote bei den Familienleistungen ist darauf zurückzuführen, dass deren quantitativ wichtigster Teil, die Familienbeihilfe, berechnet wurde.

Der Anteil der Haushalte, die beim Einkommen keinen Item-Non-Response aufweisen, liegt bei 59%. Lässt man Vermögenseinkommen (überwiegend Zinseinkommen) außer Betracht, so steigt dieser Anteil auf 70%.

Messfehler (Erfassungsfehler)

Im Gegensatz zum Zufallsfehler können Messfehler verschiedenste Gründe haben und sind in der Regel nicht quantifizierbar. Die wichtigsten Ursachen für Messfehler bei EU-SILC sind:

- der Fragebogen (z.B. missverständliche Formulierung, schlechter Ausdruck, ungeeignete Abfolge der Fragen),
- die InterviewerInnen (unpassendes Auftreten, Überspringen von Fragen, fehlerhafte Bedienung der Laptops) und
- die RespondentInnen (falsche Angaben, Proxy-Interviews).

Um solche Fehler möglichst zu vermeiden, wurde zunächst im Jahr 2002 eine Probeerhebung durchgeführt, deren Erkenntnisse in die Fragebogengestaltung und die Festlegung der Checks, aber auch in die Kooperation mit dem Erhebungsinstitut einfließen. Weiters wurden eine Reihe von Plausibilitätsprüfungen definiert und durchgeführt. Um diese möglichst effizient einsetzen und allfälligen systematischen Messfehler entgegenwirken zu können, war in der Ausschreibung die Lieferung der Daten in Tranchen vorgegeben.

Der bei der Piloterhebung verwendete Fragebogen war auf Basis eines Entwurfs von Eurostat und einer Rohübersetzung des deutschen Statistischen Bundesamtes entwickelt worden, wobei auf die nationalen Gegebenheiten Österreichs v.a. in steuerrechtlicher Hinsicht bedacht genommen werden musste. Wie bereits erwähnt, sind bei EU-SILC keine konkreten Fragen, sondern Zielvariablen vorgegeben.

Die CAPI-Erhebungstechnik erlaubt es, einen Teil der Checks bereits in die Feldarbeit zu integrieren. So erhielten die InterviewerInnen z.B. Warnmeldungen am Bildschirm, wenn ein Einkommensbetrag außerhalb eines festgelegten Intervalls lag oder wenn die Angaben zu Alter und Bildungsabschluss nicht übereinstimmten. Bei manchen Checks ist eine Implementierung in die Datenerhebung jedoch schwer möglich, sowohl aus EDV-technischen als auch aus praktischen Gründen.

Im Zuge der Datenchecks stellte sich heraus, dass einige InterviewerInnen die Frage nach dem Einkommen aus unselbständiger Tätigkeit systematisch übergingen. Dem Erhebungsinstitut wurde daraufhin mitgeteilt, dass solche Datensätze nicht akzeptiert würden. Als Reaktion darauf kam es zu einem Austausch bzw. einer Instruktion der InterviewerInnen einerseits sowie einer telefonischen Nacherhebung in den Haushalten andererseits. Von dieser Nacherhebung betroffen waren 914 Personen (9,6%) in 633 Haushalten (13,7%).

Ein anderer Erfassungsfehler war die bei einigen RespondentInnen auftretende Verwechslung von Monats- und Jahresbeträgen. Diesbezüglich besteht wohl auch Verbesserungsbedarf im Fragebogen. Weiters wurde vom Erhebungsinstitut angemerkt, dass die elektronische Erfassung von Namen und Geburtsdatum bei manchen RespondentInnen Misstrauen auslöst und so ein Grund für unvollständige oder falsche Antworten sein könnte. In EU-SILC 2004 werden der Familienname und der Geburtstag daher auf Papier erhoben und nachträglich mit den Daten verknüpft (bevor sie nach Abschluss der Vorplaus wieder gelöscht werden).

Aufarbeitungsfehler

Nach der Datenerhebung schlossen die InterviewerInnen ihre Laptops an den Telefonanschluss an und übermittelten die Daten via Internet an das Erhebungsinstitut. Dort wurden die Daten aufbereitet und ins Textformat gebracht. Nach der Übermittlung der Daten an Statistik Austria wurden diese dort ins SPSS gebracht, die Plausibilitätsprüfungen, das *data cleaning* und die Imputation durchgeführt und die Zielvariablen programmiert. Diese wurden, nach Vorliegen des authentischen Datenbestandes, ins csv-Format gebracht und via Stadium an Eurostat geliefert.

EU-SILC 2003 war für IFES die erste groß angelegte CAPI-Studie. In diesem Zusammenhang traten manche Startschwierigkeiten auf, etwa eine relativ lange Vorlaufzeit zur Implementierung

des Fragebogens in CAPI, überhitzte Laptops oder InterviewerInnen, die mit dem Notebook nicht umgehen konnten. Es muss davon ausgegangen werden, dass auch in der Aufarbeitungsphase nach der Datenerhebung Fehler passierten.

Wegen der Auslagerung der Feldarbeit und der Daten-Erstaufarbeitung liegen detaillierte Informationen über konkrete Aufarbeitungsfehler nur fallweise vor. Jene Aufarbeitungsfehler, die ersichtlich waren (z.B. abgeschnittene Felder), wurden nachträglich korrigiert.

Modellbedingte Effekte

Zum Einfluss der Imputation (von Item-Non-Response bei Einkommensfragen) auf die Genauigkeit der Ergebnisse ist anzuführen, dass bei Imputation mittels linearer Regression die Varianz des interessierenden Merkmals unterschätzt wird. Dies folgt daraus, dass bei einer Regression im Grunde genommen Mittelwerte imputiert werden (verallgemeinerte Form der *group mean imputation*), d.h. es wird zwar die Streuung zwischen verschiedenen Gruppen, nicht jedoch die Streuung innerhalb einer Gruppe berücksichtigt. Dieser Fehler fällt klarerweise umso geringer aus, je größer der Anteil der Gesamtstreuung ist, der auf die Streuung zwischen den Gruppen entfällt.

Um die Varianz des interessierenden Merkmals (des Einkommens) dennoch valide abbilden zu können, wurde bei der Imputation ein stochastischer Störterm inkludiert, d.h. die Prognosewerte aus dem Regressionsmodell wurden um ein fiktives Residuum erweitert, das in seiner Verteilung der Verteilung der geschätzten Residuen bei den beobachteten Werten entsprach.

6.3. Rechtzeitigkeit und Aktualität

Entsprechend dem bilateralen Vertrag wurden am 30. Juli 2004 die Mikrodaten via Stadium an Eurostat übermittelt. Zeitgleich wurde der standardisierte Qualitätsbericht (englisch) via E-Mail an Eurostat versendet.

Am 29. Oktober 2004 erhielt Eurostat den Abschlussbericht, in dem die auf Basis von EU-SILC berechneten Indikatoren angeführt sind.

6.4. Vergleichbarkeit

Nachdem EU-SILC 2003 in Österreich erstmalig durchgeführt wurde, kann es derzeit keine zeitliche Vergleichbarkeit geben. Ab 2004 sollte die Erstellung von Zeitreihen möglich sein. Betreffend das gesamte Haushaltseinkommen ist dabei zu berücksichtigen, dass dieses ab 2007 zusätzliche Komponenten umfassen wird (z.B. unterstellte Mieten).

EU-SILC 2003 wurde in folgenden Staaten durchgeführt: Belgien, Dänemark, Griechenland, Irland, Luxemburg, Norwegen und Österreich. Gemäß der Verordnung findet EU-SILC ab 2004 in allen EWR-Mitgliedstaaten statt, mit Ausnahme von Deutschland, Großbritannien und Niederlande sowie den neu beigetretenen Staaten, die 2005 beginnen können. Ab 2005 wollen sich weiters vier Kandidatenländer (Bulgarien, Kroatien, Rumänien, Türkei) und die Schweiz anschließen.

Die Vergleichbarkeit wird sich in den ersten Jahren auf das Nettoeinkommen beschränken, da das Bruttoeinkommen am Anfang noch nicht überall erhoben wird. Ein internationaler Vergleich der Bruttoeinkommen sollte ab 2007 möglich sein.

6.5. Kohärenz

Für Kohärenzprüfungen bieten sich mehrere Datenquellen an. Zunächst seien hier das Europäische Haushaltspanel (ECHP, letzte Welle 2001)¹¹; IFS-Datensatz) und die Konsumerhebung 1999/2000 (KE) erwähnt, ebenfalls Stichprobenerhebungen, bei denen das Netto-Haushaltseinkommen nach Komponenten erhoben worden war und deren Populationsdefinition mit der von EU-SILC übereinstimmt.

Daneben kann auf Basis der Sektorkonten aus den Volkswirtschaftlichen Gesamtrechnungen (VGR) die Summe aller Haushaltseinkommen geschätzt werden. Diese Schätzung liefert zwar keine Information über die Verteilung des Einkommens, kann aber immerhin auf bestimmte Effekte einer Stichprobenerhebung hinweisen.

Für die beiden wichtigsten Einkommenskomponenten, das Einkommen aus unselbständiger Erwerbstätigkeit und das Einkommen aus Pensionen, wurden schließlich Kohärenzprüfungen mit den Lohnsteuerdaten 2002 (LStD) vorgenommen.

Beim Einkommen aus selbständiger Erwerbstätigkeit, welches generell wesentlich schwieriger erfassbar ist als das Einkommen aus unselbständiger Erwerbstätigkeit, wären Kohärenzprüfungen mit externen Datenquellen zweifellos ein ebenso wichtiges Mittel zur Qualitätssicherung. Allerdings ist dafür nach Auffassung von Statistik Austria derzeit keine geeignete externe Datenquelle vorhanden. Die in Frage kommenden Einkommensteuerdaten sind zum einen erst nach einigen Jahren verfügbar, zum anderen muss deren Qualität auch aus inhaltlichen Gründen angezweifelt werden, da dort jenes Einkommen erfasst ist, das von den Selbständigen gegenüber der Finanzverwaltung als Grundlage für die Steuerbemessung gemeldet wird.

Vergleich mit den Sektorkonten der Volkswirtschaftlichen Gesamtrechnungen

Um einen Vergleichswert zu erhalten, wurde die Einkommensdefinition aus EU-SILC auf die Aggregate aus den VGR angewandt. Nachdem dies nicht immer unmittelbar möglich war (z.B. Herausrechnen der privaten Organisationen ohne Erwerbszweck), musste teilweise mit Überschlagszahlen gearbeitet werden.

Als VGR-Ausgangswert für das Nettohaushaltseinkommen wurde das verfügbare Einkommen des Sektors „Private Haushalte und Organisationen ohne Erwerbszweck“ herangezogen, als Vergleichswert für das Bruttohaushaltseinkommen der Saldo der Primäreinkommen bereinigt um die tatsächlichen und unterstellten Sozialbeträge der Arbeitgeber und zuzüglich der monetären Sozialleistungen.

Die Methode ergab mit den VGR-Daten einen Schätzwert für das gesamte Brutto-Haushaltseinkommen von 163 Mrd. EUR bzw. für das gesamte Netto-Haushaltseinkommen von 121 Mrd. EUR. Die korrespondierenden hochgerechneten Werte in EU-SILC 2003 betragen 130,4 Mrd. EUR bzw. 94,2 Mrd. EUR und liegen somit um 20% bzw. 22% darunter.

¹¹ Europäisches Haushaltspanel in Österreich, IFS-Interdisziplinäres Forschungszentrum Sozialwissenschaften, Version 2004.

Vergleich des gesamten Haushaltseinkommens mit den Schätzwerten aus den VGR

(Alle Beträge in Mio. EUR)

	Bruttoeinkommen		Verfügbares Einkommen	
	gesamt	ohne Vermögens-einkommen	gesamt	ohne Vermögens-einkommen
VGR-Ausgangswert aus der Sektorrechnung	176939	161885	134676	119622
Abzug der privaten Organisationen ohne Erwerbszweck	-4000	-4000	-4000	-4000
Abzug der Bevölkerung außerhalb von Privathaushalten ¹	-3026	-2763	-2287	-2023
Abzug der Entnahmen ²	-1800	-1800	-1800	-1800
Abzug der fiktiven Mieten	-5500	-5500	-5500	-5500
VGR-Schätzwert	162613	147822	121089	106299
EU-SILC-Schätzwert	130384	129613	94202	93573
Abweichung des EU-SILC-Schätzwerts vom VGR-Schätzwert				
absolut	-32229	-18209	-26887	-12726
in %	-19.8	-12.3	-22.2	-12.0

¹ Geschätzt auf Basis von Volkszählungsergebnissen auf 1.75% der Gesamtbevölkerung.
² Geschätzt auf Basis der Konsumerhebung 1999/2000 auf 1.50% der Konsumausgaben privater Haushalte.

Etwa zur Hälfte sind diese Differenzen durch Unterschiede bei den Vermögenseinkünfte bedingt: Lässt man die Vermögenseinkünfte in beiden Fällen außer Betracht, so verringert sich die Differenz jeweils auf 12%. Es ist anzunehmen, dass die Unterschiede bei den Vermögenseinkünften mehrheitlich durch Untererfassungen in der Befragung entstehen, und zwar mutmaßlich v.a. durch Unit-Non-Response.

Dass die Nettoeinkommen in EU-SILC etwas stärker von den VGR-Zahlen abweichen als die Bruttoeinkommen, könnte durch fehlende Berücksichtigung der Steuerbegünstigung von Sonderzahlungen seitens der RespondentInnen verursacht sein.

Der übrige Teil der Abweichung ist schwerer zu interpretieren. Möglich sind Fehler in EU-SILC (Coverage, Hochrechnung), Fehler in den VGR und Fehler in der Umrechnung der VGR-Aggregate auf das Einkommenskonzept von EU-SILC (diese Umrechnung kann wegen konzeptioneller Differenzen nur grob vorgenommen werden).

Vergleich mit dem Europäischen Haushaltspanel und der Konsumerhebung

Das gesamte Nettoeinkommen der privaten Haushalte betrug in EU-SILC 2003 hochgerechnet 94,2 Mrd. EUR, verglichen mit 95,8 Mrd. EUR im ECHP 2001 und 93,2 Mrd. EUR in der KE 1999/2000. Die Werte in EU-SILC sind somit etwas niedriger, wenn man die allgemeinen nominellen Einkommenszuwächse berücksichtigt. Bei der KE ist einschränkend festzuhalten, dass die dortigen Haushaltseinkommen nicht um geleistete private Transfers reduziert worden waren (in EU-SILC macht diese Reduktion 0,8 Mrd. EUR aus).

Teilweise lassen sich auch die einzelnen Komponenten miteinander vergleichen. Dabei wird deutlich, dass im ECHP vor allem das Unselbständigeneinkommen, in der KE vor allem das Selbständigeneinkommen höher ausgefallen war. Familienleistungen und Arbeitslosengelder sind dagegen in EU-SILC in Summe erwartungsgemäß höher, da Familienleistungen in der Zwischenzeit erhöht worden sind und die Arbeitslosenquote gestiegen ist.

Das geringere Gesamtniveau bei den Erwerbseinkünften kann durch eine Untererfassung von Bezügen nicht erklärt werden: Der Anteil der Personen mit Unselbständigeneinkommen ist in allen 3 Datenquellen annähernd gleich hoch, der Anteil der Personen mit Einkünfte aus selbständiger Erwerbstätigkeit in EU-SILC sogar deutlich höher. Vielmehr dürften hier selektive Ausfälle in der KE und im ECHP vorliegen, welche durch die geringe Ausschöpfungsquote (KE)

und durch die selektive Panelmortalität (ECHP)¹² bedingt sind und in Kombination mit der Gewichtung zu einer Übererfassung des aggregierten Gesamtbetrages führen. Offensichtlich führt die Vorgabe einer hohen Mindest-Antwortquote bundesweit sowie von Mindest-Antwortquoten pro Bezirk doch zu einer besseren Erfassung einkommensschwacher Haushalte in EU-SILC. (Wenn auch bestimmte Risikogruppen, etwa Nicht-EU-AusländerInnen, in EU-SILC ebenfalls nicht dem Populationsanteil entsprechend vertreten sind. So lag deren Anteil an der Gesamtbevölkerung in der Volkszählung 2001 bei 7,5%, in EU-SILC-2003 bei ca. 6%.)

Schließlich muss auch die Hochrechnung berücksichtigt werden: KE und ECHP waren dort, entsprechend den damals verfügbaren Daten, an Bevölkerungsfortschreibungen der Volkszählung 1991 angepasst worden. Diese Fortschreibungen waren, wie die Ergebnisse der Zählung 2001 zeigten, am Ende der interzensalen Periode um etwa 100.000 Personen oder 1,25% zu hoch gewesen. Deshalb kann davon ausgegangen werden, dass ein Teil der Differenzen sich aus unterschiedlichen Gewichten ergibt.

Einkommensverteilung

Das Einkommen ist in EU-SILC ungleicher verteilt als in den beiden Vergleichsdatenquellen. Während die Mittelwerte ziemlich gleich ausfallen, ist die Standardabweichung höher, was den Variationskoeffizient erhöht. Auch weicht der Koeffizient der Schiefe stärker von 0 ab (> 0 bedeutet eine rechtsschiefe Verteilung). Diese Erscheinung wird auch durch die Dezilswerte untermauert.

Verteilung des gesamten und des äquivalisierten¹ Jahres-Netto-Haushaltseinkommens

Maßzahl	Einheit	Gesamtes Netto-Haushaltseinkommen			Äquivalisiertes Netto-Haushaltseinkommen		
		EU-SILC 2003	ECHP 2001 ²	KE 1999/2000 ³	EU-SILC 2003	ECHP 2001 ²	KE 1999/2000 ³
Mittelwert	EUR	28710	28940	28740	17620	17760	17430
Standardabweichung	EUR	20090	17940	16670	10580	9350	8460
Variationskoeffizient	-	0.70	0.62	0.58	0.60	0.53	0.49
Schiefe	-	3.0	2.0	1.7	3.7	2.7	2.4
Perzentile							
10	EUR	9870	11180	10730	8150	8760	9120
20	EUR	13880	14790	14520	10470	11040	11080
25	EUR	15580	16360	16650	11510	11920	11990
30	EUR	17350	18210	18700	12270	12720	12810
40	EUR	20870	21470	22290	13940	14510	14340
50	EUR	24610	25280	26110	15710	16190	16010
60	EUR	29000	29420	29710	17450	17850	17890
70	EUR	33590	33890	34600	19740	20070	19800
75	EUR	36400	36860	37080	21160	21310	20830
80	EUR	40130	40130	39830	22880	22830	22370
90	EUR	51200	50920	49050	28660	27380	26480
Zahl der Haushalte	1,000	3281	3300	3241	3281	3300	3241

¹ Äquivalisiert mittels EU-Skala (modifizierte OECD-Skala).

² Ohne Entnahmen bei Landwirten.

³ Ohne Entnahmen und ohne fiktive Mieten.

¹² Till, M. (2001). Stichprobe und Gewichtung des Europäischen Haushaltspanels in Österreich. Diplomarbeit. Wien.

Wie bereits erwähnt ist mit großer Wahrscheinlichkeit anzunehmen, dass die Abdeckung verschiedener Einkommensgruppen in EU-SILC gleichmäßiger ist als in der KE bzw. im ECHP. Daneben können aber natürlich auch reale Veränderungen in den zwischen den Erhebungen liegenden Zeiträumen die Verteilung beeinflussen. Zumindest beim Einkommen aus unselbständiger Erwerbstätigkeit zeigten Auswertungen aus der Lohnsteuerstatistik in den vergangenen Jahren meist ein Absinken oder ein nur schwaches Wachstum der Werte in den unteren Dezilen, während in den oberen Dezilen die Beträge überdurchschnittlich zunahmen.¹³

Vergleich mit den Lohnsteuerdaten

a) Einkommen aus unselbständiger Erwerbstätigkeit

Beim Einkommen aus unselbständiger Erwerbstätigkeit zeigen sich in EU-SILC im Vergleich zu den Lohnsteuerdaten zwei markante Differenzen: Die Zahl der BezieherInnen fällt merklich geringer aus, und die Beträge streuen weniger stark um den Mittelwert. Als Vergleichswert wird stets das Brutto-2-Einkommen herangezogen, das ist das Bruttoeinkommen abzüglich der gemäß § 67 Abs. 3 bis 8 EStG mit festen Sätzen versteuerten Bezüge (das sind vor allem Abfertigungen und Urlaubsentschädigungen sowie -abfindungen).

Die hochgerechnete Zahl der ArbeitnehmerInnen ist in EU-SILC 10% geringer als in den LStD. Überdurchschnittliche Abweichungen zeigen sich bei sehr niedrigen und sehr hohen Einkommen, bei ArbeiterInnen und Angestellten und insbesondere bei nicht ganzjährig Beschäftigten. Die Zahl der ganzjährig Beschäftigten sowie der Bediensteten im öffentlichen Dienst (Beamte, Beamtinnen und Vertragsbedienstete) stimmt dagegen fast genau überein.

Unselbständig Erwerbstätige nach Bruttobezugsstufen (Brutto-Jahreseinkommen 2) bzw. Geburtsjahren und Geschlecht

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
SUMME	3294100	3669996	-375896	-10.2	1835200	1999030	-163830	-8.2	1459000	1670966	-211966	-12.7
Bruttobezugsstufe												
bis 2000 EUR	126100	300138	-174038	-58.0	47600	130766	-83166	-63.6	78500	169372	-90872	-53.7
2001 bis 4000 EUR	115900	186742	-70842	-37.9	38000	75935	-37935	-50.0	77900	110807	-32907	-29.7
4001 bis 6000 EUR	138100	159945	-21845	-13.7	45800	60529	-14729	-24.3	92200	99416	-7216	-7.3
6001 bis 8000 EUR	149900	158551	-8651	-5.5	52700	66791	-14091	-21.1	97200	91760	5440	5.9
8001 bis 10000 EUR	138800	155444	-16644	-10.7	49300	57120	-7820	-13.7	89400	98324	-8924	-9.1
10001 bis 12000 EUR	174300	154900	19400	12.5	60100	52321	7779	14.9	114200	102579	11621	11.3
12001 bis 15000 EUR	272800	233544	39256	16.8	103100	76131	26969	35.4	169700	157413	12287	7.8
15001 bis 18000 EUR	326400	242728	83672	34.5	156600	91761	64839	70.7	169800	150967	18833	12.5
18001 bis 20000 EUR	203200	179677	23523	13.1	108200	82918	25282	30.5	95000	96759	-1759	-1.8
20001 bis 25000 EUR	516000	487294	28706	5.9	336100	287523	48577	16.9	179900	199771	-19871	-9.9
25001 bis 30000 EUR	401400	421997	-20597	-4.9	276500	283450	-6950	-2.5	124900	138547	-13647	-9.9
30001 bis 35000 EUR	212400	292690	-80290	-27.4	149300	203964	-54664	-26.8	63100	88726	-25626	-28.9
35001 bis 40000 EUR	133700	192119	-58419	-30.4	99700	135118	-35418	-26.2	34000	57001	-23001	-40.4
40001 bis 50000 EUR	174500	229411	-54911	-23.9	141100	165683	-24583	-14.8	33500	63728	-30228	-47.4
50001 bis 70000 EUR	129000	171805	-42805	-24.9	103000	138239	-35239	-25.5	26000	33566	-7566	-22.5
70001 bis 100000 EUR	49600	69923	-20323	-29.1	42800	61859	-19059	-30.8	6800	8064	-1264	-15.7
über 100000 EUR	32000	29732	2268	7.6	25300	27699	-2399	-8.7	6700	2033	4667	229.6
Geburtsjahr												
bis 1937	26300	22714	3586	15.8	14900	12093	2807	23.2	11400	10621	779	7.3
1938 - 1952	499500	594076	-94576	-15.9	306600	349177	-42577	-12.2	192900	244899	-51999	-21.2
1953 - 1962	842700	911133	-68433	-7.5	460900	483420	-22520	-4.7	381800	427713	-45913	-10.7
1963 - 1972	1041700	1041895	-195	0.0	578200	564572	13628	2.4	463500	477323	-13823	-2.9
1973 - 1986	884000	1074194	-190194	-17.7	474600	573980	-99380	-17.3	409400	500214	-90814	-18.2
1987 und später	0	25984	-25984	-100.0	0	15788	-15788	-100.0	0	10196	-10196	-100.0

¹³ Vgl. etwa Statistische Nachrichten, Heft 5/2004, S. 405, sowie Heft 1/2003, S. 33.

Unselbständige nach Bundesland (Wohnort) bzw. sozialer Stellung

	EU-SILC 2003	LStSt	Abweichung	
			absolut	in %
Bundesland				
Burgenland	110400	117181	-6781	-5.8
Kärnten	220600	238328	-17728	-7.4
NÖ	607100	698797	-91697	-13.1
OÖ	551100	628272	-77172	-12.3
Szbg	227200	244871	-17671	-7.2
Stmk	464300	524159	-59859	-11.4
Tirol	271200	322065	-50865	-15.8
Vorarlberg	150000	158670	-8670	-5.5
Wien	692300	737653	-45353	-6.1
Soziale Stellung				
Lehrlinge	73300	121304	-48004	-39.6
Arbeiter	966800	1407536	-440736	-31.3
Angestellte	1346700	1590926	-244226	-15.4
Vertragsbedienstete	191900	198685	-6785	-3.4
Beamte	275100	282563	-7463	-2.6
Sonstiges	440300	68982	371318	538.3

Unselbständig Erwerbstätige nach Geschlecht bzw. Geburtsjahrgängen und Beschäftigungsdauer

	Ganzjährig beschäftigt				Nichtganzjährig beschäftigt				Unbekannt EU-SILC 2003
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		
			absolut	in %			absolut	in %	
SUMME	2713200	2635952	77248	2.9	470000	1034044	-564044	-54.5	110900
Geschlecht									
männlich	1521100	1432652	88448	6.2	259000	566378	-307378	-54.3	55000
weiblich	1192100	1203300	-11200	-0.9	211000	467666	-256666	-54.9	55800
Geburtsjahr									
bis 1972	2042500	2061635	-19135	-0.9	293500	508183	-214683	-42.2	74100
1973 und später	670800	574317	96483	16.8	176500	525861	-349361	-66.4	36700

Als Erklärungen für diese Abweichungen kommen in Betracht:

- Der unterschiedliche Coverage (so umfasst EU-SILC nur ArbeitnehmerInnen ab 16 Jahren, die mit Hauptwohnsitz in einem österreichischen Privathaushalt gemeldet waren, während die LStD alle Personen umfassen, die im entsprechenden Jahr in Österreich ein ArbeitnehmerInneneinkommen erzielten),
- Erhebungsfehler (missverständliche Formulierung der Frage, Übergehen der Frage durch die InterviewerInnen, unrichtige Antworten, Proxy-Interviews),
- der Zufallsfehler und andere Stichprobeneffekte (z.B. Auswahlrahmen, Unit-Non-Response) und
- Fehler in den LStD (denkbar sind in diesem Zusammenhang z.B. Scheinanmeldungen sowie eine falsche bzw. fehlende Zusammenführung der einzelnen Lohnzettel auf Personenebene).

Das arithmetische Mittel des Bruttoeinkommens liegt in EU-SILC um 0,7% unter dem Wert der LStD, der Median um 2,9%. Während in den unteren Perzentilen die Werte in EU-SILC deutlich höher sind als in den LStD, sind sie in der einkommensstärkeren Hälfte niedriger. So beträgt etwa das erste Quartil in EU-SILC 12.000 EUR gegenüber 9.320 EUR in den LStD, während die entsprechenden Zahlen beim dritten Quartil 28.430 EUR gegenüber 31.100 EUR lauten. Es ist somit unschwer zu erkennen, dass die Ungleichverteilung der Einkommen in der Erhebung geringer ausfällt als in der Steuerstatistik.

Diese geringere Streuung führt auch zu Abweichungen in den Einkommensniveaus bestimmter Gruppen: So ist etwa der Median des Fraueneinkommens in der Erhebung um 1,8% höher als in der Steuerstatistik, während das mediane Männereinkommen um 7,2% geringer ausfällt. Entsprechend ist in EU-SILC das Senioritätsprinzip (der Anstieg des Durchschnittseinkommens mit dem Alter) nicht so stark ausgeprägt wie in den LStSt.

Unselbständige (Schwerpunkt), Brutto-Jahreseinkommen 2

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
Mittelwert	23279	23443	-164	-0.7	27883	28473	-590	-2.1	17486	17390	96	0.6
10%	5396	2512	2884	114.8	7980	3666	4314	117.7	3700	1876	1824	97.2
20%	9840	6965	2875	41.3	14220	10228	3992	39.0	7000	4839	2161	44.7
25%	12000	9320	2680	28.8	16287	14165	2122	15.0	8540	6726	1814	27.0
30%	13800	11760	2040	17.3	18000	17660	340	1.9	10085	8512	1573	18.5
40%	16800	16509	291	1.8	21029	22028	-999	-4.5	12677	11889	788	6.6
50%	20100	20706	-606	-2.9	23500	25323	-1823	-7.2	15390	15119	271	1.8
60%	22988	24460	-1472	-6.0	26864	28760	-1896	-6.6	17800	18480	-680	-3.7
70%	26704	28571	-1867	-6.5	30156	33109	-2953	-8.9	21000	22106	-1106	-5.0
75%	28427	31103	-2676	-8.6	33041	36040	-2999	-8.3	22400	24348	-1948	-8.0
80%	31494	34298	-2804	-8.2	36940	39860	-2920	-7.3	25200	26945	-1745	-6.5
90%	43000	45164	-2164	-4.8	49000	52730	-3730	-7.1	32000	34960	-2960	-8.5

Unselbständige (Schwerpunkt), Brutto-Jahreseinkommen 2

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
Insgesamt												
Mittelwert	23279	23443	-164	-0.7	27883	28473	-590	-2.1	17486	17390	96	0.6
25%	12000	9320	2680	28.8	16287	14165	2122	15.0	8540	6726	1814	27.0
50%	20100	20706	-606	-2.9	23500	25323	-1823	-7.2	15390	15119	271	1.8
75%	28427	31103	-2676	-8.6	33041	36040	-2999	-8.3	22400	24348	-1948	-8.0
Geburtsjahr bis 1937												
Mittelwert	22223	34290	-12067	-35.2	34208	57880	-23672	-40.9	12740	13673	-933	-6.8
25%	3522	2362	1160	49.1	12000	11857	143	1.2	2770	1594	1176	73.8
50%	15000	12308	2692	21.9	18200	40045	-21845	-54.6	9975	3716	6259	168.4
75%	20400	46205	-25805	-55.8	70410	81557	-11147	-13.7	19600	14422	5178	35.9
Geburtsjahr 1938 - 1952												
Mittelwert	29542	33271	-3729	-11.2	34763	40136	-5373	-13.4	20940	23081	-2141	-9.3
25%	14000	16382	-2382	-14.5	18900	22705	-3805	-16.8	9267	10772	-1505	-14.0
50%	23000	27383	-4383	-16.0	27841	32297	-4456	-13.8	16681	19651	-2970	-15.1
75%	36400	42772	-6372	-14.9	42000	49561	-7561	-15.3	25869	31869	-6000	-18.8
Geburtsjahr 1953 - 1962												
Mittelwert	26348	28597	-2249	-7.9	30979	34981	-4002	-11.4	20731	21357	-626	-2.9
25%	15175	14805	370	2.5	19302	21926	-2624	-12.0	11168	10770	398	3.7
50%	22400	25225	-2825	-11.2	26400	30190	-3790	-12.6	17635	18665	-1030	-5.5
75%	31611	36851	-5240	-14.2	37100	42870	-5770	-13.5	26040	29264	-3224	-11.0
Geburtsjahr 1963 - 1972												
Mittelwert	23367	24035	-668	-2.8	28699	29728	-1029	-3.5	16728	17295	-567	-3.3
25%	12597	11995	602	5.0	18629	20048	-1419	-7.1	8300	7881	419	5.3
50%	21000	22487	-1487	-6.6	25000	27278	-2278	-8.4	14900	15132	-232	-1.5
75%	28925	31603	-2678	-8.5	33600	36491	-2891	-7.9	22500	24273	-1773	-7.3
Geburtsjahr 1973 - 1986												
Mittelwert	16901	13846	3055	22.1	19488	15564	3924	25.2	13917	11871	2046	17.2
25%	7833	3934	3899	99.1	9941	5218	4723	90.5	6691	2961	3730	126.0
50%	15600	11921	3679	30.9	18000	14095	3905	27.7	13116	9937	3179	32.0
75%	22400	21672	728	3.4	25480	23818	1662	7.0	18968	18899	69	0.4

Eine gesicherte Erklärung für diese geringere Streuung kann zwar nicht geliefert werden, jedoch scheinen auf Grund des Ausmaßes der Abweichungen und der Erfahrungen mit anderen Erhebungen folgende Annahmen gerechtfertigt:

- Der Non-Response (sowohl der Unit-Non-Response als auch der Item-Non-Response) dürfte in EU-SILC selektiv auftreten, d.h. dass Personen mit besonders hohem oder besonders niedrigem Einkommen häufig entweder gar nicht zur Mitarbeit bereit sind oder diese Frage nicht beantworten wollen.
- Bei Personen mit sehr hohem Einkommen ist, sofern sie die Frage beantworten, mit Falschangaben zu rechnen, die das Einkommen nach unten verzerren.
- Die Zahl der Personen mit sehr niedrigem Einkommen dürfte in den LStD überschätzt werden; dies kann durch geringfügige Einkünfte von Personen, die kein sonstiges Arbeitnehmereinkommen haben (z.B. Selbständige, Arbeitslose, Teilnahme an Truppenübungen), durch Fehler in der Zusammenführung der Lohnzettel auf Personenebene und durch Scheinanmeldungen bei der Sozialversicherung bedingt sein.
- Die geringere Streuung in EU-SILC ist teilweise durch die Imputation bedingt. Lineare Regressionsmodelle bewirken eine Unterschätzung der Varianz. Allerdings dürfte diese Unterschätzung in EU-SILC durch die Inklusion von stochastischen Störtermen in den Imputationsmodellen eher gering ausfallen.
- Daneben können Abweichungen natürlich auch durch die unterschiedlichen Coverages und durch konzeptionelle Differenzen bedingt sein.

Gemessen am Bruttoeinkommen wird das Nettoeinkommen in EU-SILC meist etwas geringer angegeben, d.h. die „Abgabenquote“ ist in der Befragung zumeist höher, was möglicherweise durch Nichtberücksichtigung der Steuerbegünstigung von Sonderzahlungen seitens der RespondentInnen verursacht ist.

Unselbständige (Schwerpunkt), Netto-Jahreseinkommen

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
Mittelwert	16208	16609	-401	-2.4	19070	19744	-674	-3.4	12606	12837	-231	-1.8
10%	4320	2204	2116	96.0	6500	3084	3416	110.8	3100	1691	1409	83.3
20%	7910	5793	2117	36.5	11045	8173	2872	35.1	5600	4200	1400	33.3
25%	9422	7630	1792	23.5	12180	10930	1250	11.4	7000	5651	1349	23.9
30%	10500	9504	996	10.5	13370	13296	74	0.6	8000	7039	961	13.7
40%	12652	12742	-90	-0.7	15400	16132	-732	-4.5	9860	9699	161	1.7
50%	14750	15405	-655	-4.3	16940	18190	-1250	-6.9	11443	11945	-502	-4.2
60%	16740	17722	-982	-5.5	19017	20355	-1338	-6.6	13340	14121	-781	-5.5
70%	19000	20297	-1297	-6.4	21000	23034	-2034	-8.8	15271	16389	-1118	-6.8
75%	20113	21859	-1746	-8.0	22720	24758	-2038	-8.2	16283	17767	-1484	-8.4
80%	21938	23740	-1802	-7.6	25000	26926	-1926	-7.2	18000	19403	-1403	-7.2
90%	28000	29852	-1852	-6.2	31690	34518	-2828	-8.2	22006	24115	-2109	-8.7

b) Einkommen aus Pensionen

Die Zahl der Bezieher von Pensionseinkünften fällt in EU-SILC hochgerechnet um 8,1% geringer aus als in den LStD, wobei die Abweichung bei den Frauen etwas stärker ist. Ein Teil dieser Differenz ist befragungstechnisch damit erklärbar, dass in EU-SILC nur Personen ab 16 Jahren interviewt werden und somit Waisenpensionen von Unter-15-Jährigen nicht als solche erfasst werden können.¹⁴ Weiters dürften auch die unterschiedlichen Coverages bei den PensionistInnen stärker ins Gewicht fallen als bei den unselbständigen Erwerbstätigen. Dies betrifft insbesondere Todesfälle und Personen außerhalb von Privathaushalten (was sich auch darin zeigt, dass die Abweichung bei Hochaltrigen sowie bei Frauen überdurchschnittlich ausfällt).

¹⁴ Das Einkommen von Unter-16-Jährigen wird im Haushaltsteil erhoben und bildet eine eigene Zielvariable. Im Jahr 2002 gab es in Österreich 26.000 Pensionsbezieher unter 19 Jahren.

PensionistInnen nach Bruttobezugsstufen (Brutto-Jahreseinkommen 2) bzw. Geburtsjahrgängen und Geschlecht

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
SUMME	1750200	1904373	-154173	-8.1	776900	838852	-61952	-7.4	973300	1065521	-92221	-8.7
Bruttobezugsstufe												
bis 2000 EUR	23300	38476	-15176	-39.4	7700	13339	-5639	-42.3	15600	25137	-9537	-37.9
2001 bis 4000 EUR	66900	81283	-14383	-17.7	13000	18193	-5193	-28.5	53900	63090	-9190	-14.6
4001 bis 6000 EUR	106900	103637	3263	3.1	17000	19308	-2308	-12.0	89900	84329	5571	6.6
6001 bis 8000 EUR	120000	133311	-13311	-10.0	20600	27172	-6572	-24.2	99400	106139	-6739	-6.3
8001 bis 10000 EUR	195100	264950	-69850	-26.4	34400	56737	-22337	-39.4	160700	208213	-47513	-22.8
10001 bis 12000 EUR	161300	160915	385	0.2	42400	49764	-7364	-14.8	118900	111151	7749	7.0
12001 bis 15000 EUR	199400	217691	-18291	-8.4	86700	91452	-4752	-5.2	112700	126239	-13539	-10.7
15001 bis 18000 EUR	208300	189619	18681	9.9	118100	98862	19238	19.5	90200	90757	-557	-0.6
18001 bis 20000 EUR	121500	115249	6251	5.4	68900	70164	-1264	-1.8	52600	45085	7515	16.7
20001 bis 25000 EUR	237000	226801	10199	4.5	156200	147651	8549	5.8	80800	79150	1650	2.1
25001 bis 30000 EUR	121200	151246	-30046	-19.9	80400	106861	-26461	-24.8	40900	44385	-3485	-7.9
30001 bis 35000 EUR	67700	73600	-5900	-8.0	45600	46952	-1352	-2.9	22100	26648	-4548	-17.1
35001 bis 40000 EUR	31500	40809	-9309	-22.8	20900	24323	-3423	-14.1	10600	16486	-5886	-35.7
40001 bis 50000 EUR	43800	43528	272	0.6	29900	26501	3399	12.8	14000	17027	-3027	-17.8
50001 bis 70000 EUR	25900	34313	-8413	-24.5	18800	23860	-5060	-21.2	7100	10453	-3353	-32.1
70001 bis 100000 EUR	12400	13414	-1014	-7.6	11200	10755	445	4.1	1300	2659	-1359	-51.1
über 100000 EUR	7800	4467	3333	74.6	5000	3812	1188	31.2	2800	655	2145	327.5
Geburtsjahr												
bis 1927	516800	597450	-80650	-13.5	184700	203638	-18938	-9.3	332100	393812	-61712	-15.7
1928-1937	552900	570304	-17404	-3.1	275500	290740	-15240	-5.2	277500	279564	-2064	-0.7
1938 - 1952	598700	620248	-21548	-3.5	278700	289052	-10352	-3.6	320000	331196	-11196	-3.4
1953 und später	81800	116371	-34571	-29.7	38000	55422	-17422	-31.4	43700	60949	-17249	-28.3

Wie auch bei den Unselbständigen ist in der Verteilung des Einkommens eine verstärkte Repräsentanz mittlerer Einkommenssegmente zu konstatieren. Die Differenzen zu den LStD sind jedoch schwächer als bei den ArbeitnehmerInnen. Der Mittelwert liegt in EU-SILC um 2,4% höher, der Median um 6,0%, das erste Quartil um 2,3%. Das dritte Quartil weicht in der Erhebung um 1,3% nach unten ab. Bedenkt man die erhebungstechnisch bedingte Unterrepräsentanz der Waisenspensionisten in EU-SILC, so erscheint ein im Schnitt etwas höheres Einkommensniveau als in der Steuerstatistik plausibel.

PensionistInnen (Schwerpunkt), Brutto-Jahreseinkommen 2

	Zusammen				Männer				Frauen			
	EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung		EU-SILC 2003	LStSt	Abweichung	
			absolut	in %			absolut	in %			absolut	in %
Mittelwert	17204	16795	409	2.4	21646	21260	386	1.8	13708	13294	414	3.1
10%	5587	5136	451	8.8	8925	8177	748	9.1	4480	4200	280	6.7
20%	8400	8099	301	3.7	12600	11177	1423	12.7	6664	6524	140	2.1
25%	9036	8833	203	2.3	14000	12602	1398	11.1	7700	7549	151	2.0
30%	10080	9132	948	10.4	15400	13990	1410	10.1	8400	8394	6	0.1
40%	12458	11403	1055	9.3	16800	16646	154	0.9	9394	8833	561	6.4
50%	14700	13869	831	6.0	19246	19012	234	1.2	11018	10461	557	5.3
60%	16927	16745	182	1.1	21000	21455	-455	-2.1	12880	12418	462	3.7
70%	20272	19937	335	1.7	23418	24374	-956	-3.9	15739	14987	752	5.0
75%	21500	21778	-278	-1.3	25747	25995	-248	-1.0	16969	16607	362	2.2
80%	23100	23968	-868	-3.6	27860	27661	199	0.7	18911	18643	268	1.4
90%	30100	30095	5	0.0	35000	34594	406	1.2	24147	25028	-881	-3.5

c) Abschätzung der Auswirkung der unterschiedlichen Coverages

Nachdem die EU-SILC-Schätzwerte betreffend die Zahl der Arbeitnehmer sowie der Pensionisten mit 10% bzw. 8% doch relativ stark von den LStD abweichen, war es von Interesse herauszufinden, welche Anteile dieser Differenzen mutmaßlich durch die unterschiedlichen Coverages bedingt sind. Dazu ist anzumerken, dass darüber in Ermangelung entsprechender Datenquellen keine exakten Angaben gemacht werden können. Es ist jedoch möglich, auf Basis von Schätzungen, die auf externen Datenquellen beruhen und teilweise durch Annahmen ergänzt wurden, definitorisch vergleichbare Größen einander auf Makroebene gegenüberzustellen.

Zahl der Arbeitnehmer: Wie aus der untenstehenden Tabelle entnommen werden kann, verringert sich der Vergleichswert der LStD um knapp 100.000 Personen, wenn die Definitionen

von EU-SILC zu Grunde gelegt werden. Die Abweichung des EU-SILC-Schätzers von diesem Wert ist allerdings mit 283.000 Personen bzw. 7,9% nach wie vor hoch, d.h. der unterschiedliche Abdeckungsbereich erklärt nur ca. ein Viertel der rohen Gesamtdifferenz. Diese verbleibenden 7,9% sind auch durch den Zufallsfehler nicht erklärbar, da der adaptierte LStD-Wert außerhalb des auf Basis der EU-SILC-Daten geschätzten Konfidenzintervalls für den wahren Parameter liegt. Somit müssen hier systematische Fehler vorliegen, wobei aus den Daten nicht hervorgeht, ob die systematischen Fehler in der Erhebung, in den Administrativdaten oder in beiden vorkommen. Interessant ist jedenfalls, dass die Abweichung bei den Frauen mit 11,2% sehr viel stärker ausfällt als bei den Männern mit 5,1%.

Zahl der Arbeitnehmer in EU-SILC und in den Lohnsteuerdaten (LStD): Schätzung der konzeptionell bedingten Abweichung

	Zahl der Arbeitnehmer in 1,000			Anmerkung
	Zusammen	Männer	Frauen	
Wert aus den LStD 2002	3670.0	1999.0	1671.0	a
PLUS				
Grenzgänger ins Ausland	27.3	19.6	7.7	b
Zugezogene seit Jänner 2003	29.8	18.2	11.6	c
MINUS				
Personen, die nach 1986 geboren sind	26.0	15.8	10.2	
Personen außerhalb von Privathaushalten	46.7	36.7	10.0	d
Weggezogene seit Jänner 2002	67.0	43.0	24.0	c
Gestorbene seit Jänner 2002	9.6	7.1	2.5	e
<hr/>				
Adaptierter Wert aus den LStD	3577.8	1934.2	1643.6	
EU-SILC Schätzwert	3294.2	1835.2	1459.0	
Geschätzter EU-SILC-Standardfehler	38.2	33.5	29.9	f
95%-Konfidenzintervall des wahren Werts				
untere Schranke	3219.3	1769.5	1400.4	
obere Schranke	3369.1	1900.9	1517.6	

a Personen mit Aktivbezug (unselbständige Erwerbstätigkeit) und Wohnsitz in Österreich.

b Personen, die laut Einkommensteuerstatistik 2002 das Formular E16 ausgefüllt haben.

c Wanderungsbewegungen aus der MIGSTAT gegliedert nach einjährigen Altersgruppen und Geschlecht, multipliziert mit den zellenspezifischen Arbeitnehmerquoten aus der Volkszählung 2001.

d Erwerbspersonen in Anstaltshaushalten (ISIS-Segment B4M) und Gemeinschaftsunterkünften (ISIS-Segment E2A), multipliziert mit einem unterstellten Arbeitnehmeranteil von 90%.

e Gestorbene aus dem POPREG gegliedert nach einjährigen Altersgruppen und Geschlecht, multipliziert mit den zellenspezifischen Arbeitnehmerquoten aus der Volkszählung 2001.

f Geschätzt mittels Bootstrapping.

Zahl der Pensionisten: Dieselben Adaptierungen wie bei den Arbeitnehmern wurden auch bei den Pensionisten vorgenommen, bis auf die Ausnahme, dass hier zu den LStD nicht die Grenzgänger hinzuaddiert werden mussten, sondern die Pensionsbezieher mit ausschließlich ausländischer Pension. Letztere Zahl ist wegen diverser bilateraler Besteuerungsabkommen auch in den Einkommensteuerdaten nur mangelhaft eruiert. Jedenfalls zeigt sich, dass bei Berücksichtigung des geschätzten Einflusses der unterschiedlichen Coverages die EU-SILC-Schätzwerte verhältnismäßig gut mit den LStD übereinstimmen: Aus der rohen Unterschätzung

von 8% wird eine Überschätzung von 2%. Diese Differenz liegt bereits innerhalb des auf Basis der EU-SILC-Daten geschätzten Konfidenzintervalls für den wahren Parameter, d.h. sie ist durch den Zufallsfehler erklärbar und die Nullhypothese, wonach der EU-SILC-Wert dem bereinigten LStD-Wert entspricht, kann bei einer Irrtumswahrscheinlichkeit von 5% nicht verworfen werden.

Zahl der Pensionisten in EU-SILC und in den Lohnsteuerdaten (LStD): Schätzung der konzeptionell bedingten Abweichung

	Zahl der Pensionisten in 1,000			Anmerkung
	Zusammen	Männer	Frauen	
Wert aus den LStD 2002	1904.4	838.9	1065.5	a
PLUS				
Personen mit nur ausländischer Pension	10.0	4.4	5.6	b
Zugezogene seit Jänner 2003	2.5	1.3	1.2	c
MINUS				
Personen, die nach 1986 geboren sind	15.5	7.9	7.6	
Personen außerhalb von Privathaushalten	78.5	29.0	49.5	d
Weggezogene seit Jänner 2002	14.6	7.4	7.2	c
Gestorbene seit Jänner 2002	96.6	43.6	53.0	e
<hr/>				
Adaptierter Wert aus den LStD	1711.7	756.7	955.0	
EU-SILC Schätzwert	1750.2	776.9	973.3	
Geschätzter EU-SILC-Standardfehler	30.5	21.8	24.4	f
95%-Konfidenzintervall des wahren Werts				
untere Schranke	1690.4	734.2	925.5	
obere Schranke	1810.0	819.6	1021.1	

a Personen mit Pensionsbezug und Wohnsitz in Österreich.

b Diese Größe kann wegen fehlender Daten nicht geschätzt, sondern nur angenommen werden.

c Wanderungsbewegungen aus der MIGSTAT gegliedert nach einjährigen Altersgruppen und Geschlecht, multipliziert mit den zellenspezifischen Pensionistenquoten aus der Volkszählung 2001.

d Berufslose Einkommensempfänger in Anstaltshaushalten (ISIS-Segment B4M) und Gemeinschaftsunterkünften (ISIS-Segment E2A), multipliziert mit einem unterstellten Pensionistenanteil von 90%.

e Gestorbene aus dem POPREG gegliedert nach einjährigen Altersgruppen und Geschlecht, multipliziert mit den zellenspezifischen Pensionistenquoten aus der Volkszählung 2001.

f Geschätzt mittels Bootstrapping.

Hinweis auf ergänzende Dokumentationen/Publikationen

[Strukturindikatoren der Europäischen Union](#) (Beschreibung und Daten).