

Helga Christians / Heike Wirth

**Mikrozensus-Regionalfile:  
Hochrechnung, Stichprobenfehler und  
Anonymisierung**

(unter Mitarbeit von:  
Baris Bayindir und Reinhard Schunck)

Düsseldorf, Mannheim

November 2009

# **Mikrozensus-Regionalfile: Hochrechnung, Stichprobenfehler und Anonymisierung**

<b>1</b>	<b>Einleitung .....</b>	<b>3</b>
<b>2</b>	<b>Hochrechnung im Mikrozensus .....</b>	<b>4</b>
2.1	Hochrechnungsverfahren auf Ebene der Anpassungsschichten	4
2.2	Hochrechnungsverfahren für das Mikrozensus-Regionalfile	5
2.2.1	Auswahl der regionalen Einheit .....	5
2.2.2	Bildung der Hochrechnungsfaktoren auf MZKR-Ebene .....	6
2.3	Ergebnisse auf Basis des Hochrechnungsverfahrens auf MZKR-Ebene	6
2.3.1	Vergleich der hochgerechneten Ergebnisse mit der Bevölkerungfortschreibung für die Merkmale Staatsangehörigkeit und Geschlecht .....	7
	Quelle: Mikrozensus 2000.....	8
2.3.2	Konsistenzprüfung der Ergebnisse der Hochrechnung auf MZKR-Ebene und Anpassungsschichtebene .....	8
2.3.3	Zusammenfassung der Ergebnisse .....	14
<b>3</b>	<b>Stichprobenfehler .....</b>	<b>16</b>
<b>4</b>	<b>Anonymisierungsmaßnahmen .....</b>	<b>20</b>
4.1	Die Anonymisierungsempfehlungen	20
4.2	Beruf und Wirtschaftszweig	22
4.3	Alter	26
4.4	Nationalität	26
<b>5</b>	<b>Fazit .....</b>	<b>29</b>
<b>6</b>	<b>Literatur: .....</b>	<b>31</b>

**Anhänge:**

Anhang 1: Häufigkeiten von ausgewählten Anpassungsschichten aus Bayern	32
Anhang 2: Anonymisierung des Wirtschaftszweigs: Darstellung der Anteile nach Anonymisierung auf Bundes- und MZKR-Ebene	33

Folgende weitere Informationen zum Mikrozensus-Regionalfile befinden sich auf der CD zum Datenfile:

- Dokumentation zum Stichprobenfehler: Exemplarische Fehlerrechnungen für ausgewählte Merkmale auf Mikrozensuskreisebene.
- Umsetzung der Anonymisierung: Umsetzung des Anonymisierungskonzepts des Mikrozensus-Regionalfiles als SUF für die Off-Site-Nutzung für das Jahr 2000.
- Vergleich Originalmaterial und Stichprobe: Vergleich der Ergebnisse der Originaldaten und der 85% Stichprobe des Regionalfiles auf Bundeslandsebene.
- MZ-Regionalfile - Schlüsselverzeichnis MZ 2000: Beinhaltet eine Auflistung aller im Regionalfile verfügbaren Merkmale. Darüber hinaus ist hier merkmalspezifisch dokumentiert welche Kategorien anonymisierungsbedingt zusammengefasst wurden.
- Zusammenfassung MZKR: Dokumentation der Zusammenfassung der Kreise in Mikrozensuskreisregionen (MZKR).
- Regionale Schichtung nach MZ-Auswahlplan (Stand 2000): Aufstellung zu den regionalen Schichten im MZ.
- Raumordnungsregionen (Stand 2000): Aufstellung der Raumordnungsregionen für das Jahr.
- Berufsklassifikationen 1992: Klassifikation der Berufe, Ausgabe 1992
- Wirtschaftszweigklassifikation 1993: Klassifikation der Wirtschaftszweige, Ausgabe 1993. Fassung für den Mikrozensus

## 1 Einleitung

Die Bedeutung regionaler Disparitäten für die Lebenschancen unterschiedlicher gesellschaftlicher Gruppen rückt zunehmend in den Vordergrund wissenschaftlicher Untersuchungen. Der Mikrozensus stellt mit seinem hohen Auswahlsatz, der thematischen Breite und zeitlichen Kontinuität in diesem Zusammenhang eine wichtige Datenquelle dar. Um diese Analysen zu ermöglichen wird der Wissenschaft ein Mikrozensus-Regionalfiler als Scientific Use File (SUF) für die Off-Site-Nutzung angeboten. Dieses SUF ermöglicht Analysen auf Ebene so genannter Mikrozensuskreisregionen (MZKR). Diese MZKR sind regionale Einheiten, die mindestens 100.000 Einwohner umfassen. Die Abgrenzungen der MZKR orientieren sich an den Raumordnungsregionen bzw. siedlungsstrukturellen Kreistypen<sup>1</sup> und bieten vielfältige Möglichkeiten zu weiteren forschungsorientierten Aggregationen. So ist es z.B. möglich, sie zu Arbeitsmarktregionen zusammenzufassen oder eigene Aggregationen für spezifische Fragestellungen vorzunehmen.

Für das Mikrozensus-Regionalfiler wurden Hochrechnungsfaktoren generiert, die eine Hochrechnung auf Ebene der MZKR ermöglichen. Im 2. Kapitel des vorliegenden Papiers wird die Hochrechnungsmethode, die sich an dem bisherigem Hochrechnungsverfahren im Mikrozensus orientiert, näher erläutert und es wird die Qualität der Ergebnisse dargestellt. Mit der Hochrechnung sollen Befragungsausfälle kompensiert und die Ergebnisse an die Zahlen der Bevölkerungsfortschreibung angepasst werden. Zudem wird im Vergleich zu einer freien Hochrechnung eine Präzisierung der Ergebnisse sowie die Verringerung des Standardfehlers erwartet. Das vorliegende Papier behandelt die Mikrozensusjahrgänge von 1990 bis 2004.

Das 3. Kapitel beschäftigt sich mit dem Ausmaß des Stichprobenfehlers im Mikrozensus-Regionalfiler. Der Stichprobenfehler fällt umso größer aus, je kleiner die Fallzahlen für bestimmte Merkmalskombinationen in den betrachteten räumlichen Einheiten werden. Im Anschluss an die Darstellungen zum Stichprobenfehler werden Empfehlungen für die Interpretation von Ergebnissen insbesondere bei Punktschätzungen gegeben.

Abschließend werden im 4. Kapitel die Anonymisierungsmaßnahmen für das Mikrozensus-Regionalfiler als SUF für die Off-Site-Nutzung dargestellt. Das Ziel ist es hierbei, die faktische Anonymität der Daten mit einem möglichst geringen Informationsverlust für die Datenanalyse zu realisieren (vgl. Müller u.a., 1991).

Das Konzept wurde gemeinsam vom German Microdata Lab des GESIS Leibniz-Institut für Sozialwissenschaften und dem Forschungsdatenzentrum der Statistischen Landesämter erarbeitet. Die Umsetzung des Konzepts wurde zunächst für das Jahr 2000 durchgeführt. Neben der Bereitstellung eines SUF für die Off-Site-Nutzung können die Daten am Gastwissenschaftlerarbeitsplatz oder mittels kontrollierter Datenfernverarbeitung genutzt werden.

---

<sup>1</sup> Eine Aufstellung der MZKR befindet sich im PDF-Dokument: Zusammenfassung\_MZKR

## 2 Hochrechnung im Mikrozensus

Ziel der Hochrechnung im Mikrozensus ist es, durch den Ausgleich der bekannten Befragungsausfälle und die Randanpassung an Zahlen der Bevölkerungsfortschreibung eine Verbesserung der Ergebnisse des Mikrozensus zu erreichen. Im Folgenden wird zunächst das bestehende Hochrechnungsverfahren auf Ebene der Anpassungsschicht erläutert (Abschnitt 2.1). Es dient als Ausgangspunkt für die Entwicklung eines Hochrechnungsverfahrens für das Mikrozensus-Regionalfiler. Im Abschnitt 2.2 wird das Hochrechnungsverfahren für das Mikrozensus-Regionalfiler vorgestellt. Abschnitt 2.3 beschäftigt sich mit den Ergebnissen, die bei der Anwendung des regionalfiler-spezifischen Hochrechnungsverfahrens erzielt werden.

### 2.1 Hochrechnungsverfahren auf Ebene der Anpassungsschichten

Hochrechnungsfaktoren stehen im Mikrozensus bisher auf Ebene der Anpassungsschichten zur Verfügung. Anpassungsschichten sind Zusammenfassungen von Kreisen, die im Durchschnitt 500.000 Einwohner umfassen.

Die Hochrechnungsfaktoren der Anpassungsschichten bestehen aus einem Kompensationsfaktor, der die bekannten Ausfälle im Datensatz ausgleicht und einem Anpassungsfaktor, der die Ergebnisse des Mikrozensus an die Zahlen der Bevölkerungsfortschreibung des Vorjahres bzw. an die interpolierten Werte zum Zeitpunkt der Mikrozensus-Berichtswoche angleicht. Der *Kompensationsfaktor* wird auf Ebene der regionalen Untergruppen ermittelt. Regionale Untergruppen sind Regionaleinheiten von ca. 100.000 Einwohnern und können aus Teilen von Kreisen, aus einzelnen Kreisen oder aber auch aus mehreren Kreisen bestehen. Die festgestellten Ausfälle werden in verschiedene Typen eingeteilt. Relevante Merkmale für die Einteilung in die einzelnen Kompensationstypen sind Haushaltsgröße und Staatsangehörigkeit. Für deutsche Haushalte wird nach Möglichkeit festgestellt, ob es sich bei der Wohnung um eine Haupt- oder Nebenwohnung handelt und für 1-Personenhaushalte werden, soweit möglich, das Geschlecht und das Alter der Person festgestellt. In einem mehrstufigen Verfahren werden die bekannten Ausfälle je nach Informationsgrad am Ende in insgesamt 19 Kompensationstypen eingeteilt (12 für Einpersonenhaushalte, 6 für Mehrpersonenhaushalte und ein Typ für die Gemeinschaftsunterkünfte). Für diese Typen werden dann auf Ebene der regionalen Untergruppen die entsprechenden Kompensationsfaktoren ermittelt in dem sie mit den Istzahlen der jeweils gleichen Haushaltsgruppe ins Verhältnis gesetzt werden. Die Berechnung der *Anpassungsfaktoren* erfolgt pro Anpassungsschicht jeweils für Ausländer und Deutsche getrennt nach dem Geschlecht. Die Sollzahlen werden der Bevölkerungsfortschreibung – unterteilt in diese 4 Kategorien – entnommen. Dazu werden die Zahlen zum Ende des Vorjahres linear interpoliert, um so die Bevölkerungszahlen zum Zeitpunkt der Mikrozensususerhebung zu schätzen (Statistisches Bundesamt 1997:17 u. 43). Diese Zahlen werden als Sollzahlen in den jeweiligen Kategorien mit den Istzahlen der Mikrozensususerhebung nach durchgeführter Kompensation ins Verhältnis gesetzt. Dabei wird vorab die Zahl der Soldaten und Wehrpflichtigen von der Soll- und Istzahl der deutschen Männer abgezogen. Sofern in einer Anpassungsschicht die Istzahl in einer der vier Kategorien kleiner 50 ist, wird diese mit einer der anderen Kategorien zusammengefasst und ein gemeinsamer Anpassungsfaktor für beide Personengruppen ermittelt: Weist beispielsweise die Kategorie „ausländische Frauen“ nur 39 Fälle auf, wird sie mit den „deutschen Frauen“ zusammengefasst. Der entsprechende Anpassungsfaktor bezieht sich dann auf

deutsche und ausländische Frauen. Für Soldaten und Wehrdienstleistende werden separate Anpassungsfaktoren auf Regierungsbezirksebene ermittelt.

Das Hochrechnungsverfahren beschränkt sich zunächst auf die Personen am Ort der Hauptwohnung, da sich die Bevölkerungsfortschreibung lediglich auf diesen Personenkreis bezieht. Den Personen am Ort der Nebenwohnung werden die Anpassungsfaktoren ihrer jeweiligen Bezugsgruppe (Deutsche/Ausländer unterschieden nach Geschlecht sowie der Gruppe der Soldaten und Wehrdienstleistenden) zugeordnet.

Der *letztendliche Hochrechnungsfaktor* der einzelnen Personen ergibt sich dann aus dem Produkt von Anpassungsfaktor und Kompensationsfaktor. Aus den Hochrechnungsfaktoren auf Personenebene werden weiterhin entsprechende Hochrechnungsfaktoren für die Haushalte generiert. Der Hochrechnungsfaktor auf Haushaltsebene ergibt sich aus dem arithmetischen Mittel der Hochrechnungsfaktoren der zum Haushalt gehörenden Personen.<sup>2</sup>

## **2.2 Hochrechnungsverfahren für das Mikrozensus-Regionalfile**

Ziel ist es, für das Mikrozensus-Regionalfile möglichst differenzierte Hochrechnungsfaktoren zur Verfügung zu stellen, um eine gute Anpassung an die Zahlen der Bevölkerungsfortschreibung in den Kategorien ausländische und deutsche Frauen bzw. Männer für die einzelnen Kreise zu erreichen. Als regionale Einheit für die Entwicklung von Hochrechnungsfaktoren kommen die Kreise und die MZKR in Betracht. Aus den in Abschnitt 2.2.1 dargestellten Gründen werden die Hochrechnungsfaktoren auf Ebene der MZKR entwickelt. In Abschnitt 2.2.2 wird das für das Mikrozensus-Regionalfile angepasste Hochrechnungsverfahren beschrieben.

### **2.2.1 Auswahl der regionalen Einheit**

Nahe liegend erscheint die Bildung von Hochrechnungsfaktoren für das Regionalfile auf Kreisebene, um Verzerrungen, die bei der Bereitstellung gemeinsamer Hochrechnungsfaktoren für mehrere Kreise entstehen können, zu vermeiden. Die Arbeiten am Hochrechnungsverfahren haben jedoch gezeigt, dass in Kreisen mit weniger als 100.000 Einwohner und einem geringen Ausländeranteil, keine gesonderten Hochrechnungsfaktoren für ausländische und deutsche Frauen bzw. Männer mehr ausgewiesen werden können, da die Fallzahl des Öfteren eine kritische Grenze von unter 10 Fällen erreicht. Im Ergebnis würde diese Herangehensweise daher zu erheblichen Verzerrungen in der Darstellung der Bevölkerungsstruktur auf Kreisebene aber auch für darauf aufbauende Aggregationen führen.

Da das Hochrechnungsverfahren so angelegt sein sollte, dass für die gewählten regionalen Gliederungen möglichst differenzierte Hochrechnungsfaktoren nach Deutschen und Ausländern gebildet werden, mit denen für die unterschiedlichsten Aggregationen gute Ergebnisse erzielt werden, bleibt als Lösung die Entwicklung eines Hochrechnungsverfahrens, das auf die MZKR aufsetzt. Hier können

---

<sup>2</sup> Vgl. allgemein zur Hochrechnung im Mikrozensus: Heidenreich, 1994 sowie zur genauen Programmroutine des Statistischen Bundesamtes: Statistisches Bundesamt 1997 Abschnitt 2.12, S. 32.

für die Mehrzahl der MZKR getrennte Hochrechnungsfaktoren für ausländische und deutsche Frauen und Männer ausgewiesen werden.

### 2.2.2 Bildung der Hochrechnungsfaktoren auf MZKR-Ebene

Das Hochrechnungsverfahren auf MZKR-Ebene orientiert sich an dem bisherigen Verfahren für den Mikrozensus und sichert damit eine weitgehende Konsistenz der Hochrechnung auf den unterschiedlichen regionalen Ebenen. Die MZKR-Hochrechnungsfaktoren wurden wie folgt ermittelt:

- Auf die Neuberechnung des *Kompensationsfaktors* wird verzichtet. Er wird standardmäßig auf der Ebene der so genannten regionalen Untergruppen berechnet. Der Kompensationsfaktor liegt somit bereits in einer tiefen regionalen Untergliederung vor.
- Für die Ermittlung des *Anpassungsfaktors* für Ausländer und Deutsche – unterschieden nach dem Geschlecht – werden die Zahlen der Bevölkerungsfortschreibung zum Ende des jeweiligen Vorjahres der Mikrozensus Erhebung als Sollzahlen verwendet. Auf eine Interpolation der Werte wird verzichtet. Als Istzahlen für die Berechnung der Anpassungsfaktoren werden die Häufigkeiten des Mikrozensus auf MZKR-Ebene nach Kompensation der Ausfälle genutzt. Von der Ist- und der Sollzahl der männlichen Deutschen wird im weiteren Verlauf – analog zur Programmroutine des Statistischen Bundesamtes – die Istzahl der Soldaten und Wehrpflichtigen abgezogen. Im Unterschied zu der ursprünglichen Programmroutine des Statistischen Bundesamtes wird für die Zusammenfassung von Kategorien zur Berechnung gesonderter Hochrechnungsfaktoren für ausländische Männer und Frauen, der Grenzwert von bisher 50 Fällen pro Zelle auf einen Grenzwert von 10 Fällen pro Kategorie festgelegt. Ein Festhalten an einem Grenzwert von 50 Fällen würde bedeuten, dass für viele MZKR keine gesonderten Hochrechnungsfaktoren für ausländische Männer und Frauen gebildet werden können. Dieses hätte nicht nur Auswirkungen auf die Ergebnisse auf Kreisebene, sondern würde auch die Ergebnisse auf höherer regionaler Aggregationsebene erheblich beeinflussen. Ein vollständiger Verzicht auf einen Grenzwert ist - nach den vorliegenden Analysen – ebenfalls nicht sinnvoll, da die Hochrechnungsfaktoren für ausländische Personen in MZKR mit niedrigem Ausländeranteilen sehr hoch werden, so dass dieses wiederum zu Verzerrungen bei inhaltlichen Analysen führen können.
- Abschließend wird der neu ermittelte *Anpassungsfaktor* mit dem jeweiligen *Kompensationsfaktor* der Person multipliziert.
- Für Soldaten und Wehrdienstleistende wird der bisherige Hochrechnungsfaktor beibehalten.
- Der *Haushaltshochrechnungsfaktor* ergibt sich als arithmetisches Mittel aus den jeweiligen Hochrechnungsfaktoren der Haushaltsmitglieder.

## 2.3 Ergebnisse auf Basis des Hochrechnungsverfahrens auf MZKR-Ebene

Im Folgenden wird dargestellt, wie sich das Hochrechnungsverfahren auf MZKR-Ebene auf die Ergebnisse im Mikrozensus auswirkt. Die Leistungsfähigkeit des Verfahrens wird hierbei anhand von zwei Aspekten beurteilt:

(1) Mit den Hochrechnungsfaktoren soll eine Anpassung an die Zahlen der Bevölkerungsfortschreibung in den Kategorien ausländische und deutsche Männer und Frauen für die einzelnen MZKR er-

reicht werden. Hierzu werden die Zahlen der Bevölkerungsfortschreibung mit den Ergebnissen der Hochrechnung auf MZKR-Ebene verglichen.

(2) Weiterhin sollten die auf Basis der MZKR hochgerechneten Ergebnisse weitgehend konsistent mit den Ergebnissen des bisherigen Hochrechnungsverfahrens auf Anpassungsschichtebene sein. Hier wird das Merkmal Staatsangehörigkeit und die Anzahl der Haushalte betrachtet. Verglichen werden zudem die sich ergebenden Ergebnisse auf Basis der beiden Hochrechnungsverfahren für ein Merkmal, an das nicht angepasst wurde. Hier wurde das Merkmal „höchster beruflicher Abschluss“ gewählt.

Das Hochrechnungsverfahren wurde zunächst für das Bundesland NRW für das Jahr 2000 umgesetzt und dann auf die weiteren Bundesländer übertragen. Die dargestellten Ergebnisse beziehen sich deshalb vorrangig auf NRW und werden – soweit notwendig – durch Ergebnisse für andere Bundesländer bzw. das gesamte Bundesgebiet ergänzt.

### 2.3.1 Vergleich der hochgerechneten Ergebnisse mit der Bevölkerungsfortschreibung für die Merkmale Staatsangehörigkeit und Geschlecht

Mit den MZKR-Hochrechnungsfaktoren wird in NRW für die Kategorien deutsche Männer, deutsche Frauen, ausländische Männer und ausländische Frauen in fast allen Kreisen eine völlige Übereinstimmung mit den Zahlen der Bevölkerungsfortschreibung des Vorjahres erreicht. Insgesamt wird die Einwohnerzahl von NRW um 13 Personen unterschätzt. Dieses resultiert aus kleineren Differenzen, die in drei Kreisen auftreten. Auch für Personengruppen, die im Mikrozensus mit einer sehr geringen Fallzahl auftreten – dieses gilt insbesondere für die ausländischen Männer und Frauen – wird eine sehr gute Anpassung erreicht. Dies liegt daran, dass für NRW in keinem der MZKR die Mindestgrenze von 10 Fällen pro Anpassungsklasse unterschritten wird.

Vergleichbare Ergebnisse für die Anpassung an die Zahlen der Bevölkerungsfortschreibung ergeben sich in den meisten MZKR auch für die weiteren westdeutschen Bundesländer. Abweichungen zu den Eckwerten der laufenden Bevölkerungsfortschreibung treten nur in solchen MZKR auf, in denen die Besetzung in den Anpassungsklassen der ausländischen Männer oder Frauen unter 10 Fällen liegt und deshalb nur ein gemeinsamer Hochrechnungsfaktor für deutsche und ausländische Männer bzw. deutsche und ausländische Frauen ausgewiesen werden kann. Dies betrifft vorwiegend die ostdeutschen Bundesländer, bei denen in 65 von 79 MZKR die Mindestfallzahl von 10 für ausländische Männer oder Frauen nicht erfüllt ist (vgl. Tabelle 1). Bei der Hochrechnung macht sich dies insofern bemerkbar, als der Anteil dieser Personengruppen an der Bevölkerung um bis zu 90% unterschätzt wird<sup>3</sup>. Allerdings zeigt Tabelle 1 auch, dass es sich hier um kein spezifisches Problem der MZKR handelt, vielmehr tritt dieses Problem bereits auf der Ebene der Anpassungsschichten auf: In allen Anpassungsschichten der ostdeutschen Bundesländer wird die Mindestbesetzung von 50 ausländischen Männern und/oder Frauen nicht erreicht und dementsprechend auch keine gesonderten Hochrech-

---

<sup>3</sup> In einigen MZKR, in denen kein/e einzige/r Ausländer/in befragt wurden, beträgt die Abweichung zur Bevölkerungsfortschreibung zwangsläufig 100%.



nungsfaktoren gebildet. Deshalb wird auch auf der Ebene der Anpassungsschichten die Anzahl der Ausländer im Vergleich zur Bevölkerungsfortschreibung um bis zu 86% unterschätzt.

Tabelle 1: Anzahl der MZKR und Anpassungsschichten (ANPS) nach Bundesländern, in denen die Zellbesetzung bei <10 Fällen bzw. <50 Fällen für ausländische Männer und/oder Frauen liegt

Bundesland	Mikrozensuskreisregionen (MZKR)		Anpassungsschichten (ANPS)	
	Anzahl MZKR	Anzahl der MZKR mit <10 Fällen in einer Anpassungsklasse	Anzahl ANPS	Anzahl der ANPS mit <50 Fällen in einer Anpassungsklasse
Schleswig-Holstein	13	2	3	0
Hamburg	1	0	1	0
Niedersachsen	39	7	11	0
Bremen	2	0	2	1
Nordrhein-Westfalen	54	0	34	0
Hessen	25	2	9	0
Rheinland-Pfalz	24	0	8	1
Baden-Württemberg	43	0	13	0
Bayern	63	7	22	0
Saarland	5	0	1	0
Berlin	1	0	2	0
Brandenburg	15	13	5	5
Mecklenburg-Vorpommern	11	10	4	4
Sachsen	23	17	9	9
Sachsen-Anhalt	14	11	5	5
Thüringen	16	14	4	4
<b>Gesamt</b>	<b>349</b>	<b>83</b>	<b>133</b>	<b>29</b>

Quelle: Mikrozensus 2000

### 2.3.2 Konsistenzprüfung der Ergebnisse der Hochrechnung auf MZKR-Ebene und Anpassungsschichtebene

Insgesamt lässt sich eine hohe Übereinstimmung der Ergebnisse mit den Anpassungsschicht- und MZKR-Hochrechnungsfaktoren feststellen. Vergleicht man die Ergebnisse, die mit den MZKR-Hochrechnungsfaktoren ermittelt wurden mit jenen auf Grundlage der Hochrechnungsfaktoren auf Anpassungsschichtebene, so erhält man für NRW zum größten Teil übereinstimmende Ergebnisse. Bei Verwendung der MZKR-Hochrechnungsfaktoren erhöht sich im Vergleich zu den Hochrechnungsfaktoren der Anpassungsschicht die Zahl der Einwohner pro Anpassungsschicht leicht (ca. 0,02-0,15%). Dabei ist nahezu unerheblich, ob es sich bei der Anpassungsschicht um einen einzelnen Kreis handelt oder eine Zusammenfassung von mehreren Kreisen. Für NRW gesamt ergibt sich eine geringe Differenz von 11.118 Personen. Bei der Bevölkerungszahl des gesamten Bundesgebietes kommt es zu keinen nennenswerten Abweichungen. Bezogen auf die Anpassungsschichten sind die

mit den beiden Hochrechnungsfaktoren berechneten Einwohnerzahlen sowohl für die neuen als auch die alten Bundesländer vergleichbar. Die Differenzen zwischen den beiden Hochrechnungsverfahren betragen in den meisten Fällen weniger als 0,5 %.

*Staatsangehörigkeit:*

Betrachtet man zunächst für NRW die Häufigkeiten getrennt nach deutschen und ausländischen Männern bzw. Frauen, so lässt sich feststellen, dass die Abweichungen bei den deutschen Frauen und Männern fast immer unter 0,1% liegen. Auch bei der ausländischen Bevölkerung kommt es in NRW zu keinen größeren Abweichungen. Insgesamt ist die Zahl der ausländischen Frauen und Männern mit dem MZKR-Hochrechnungsfaktor etwas höher als mit dem Hochrechnungsfaktor der Anpassungsschicht. Jedoch beträgt die Abweichung in keinem Fall mehr als 158 Personen, das ist eine Abweichung von unter 1%. Auch bei den Anpassungsschichten, die mehrere kleine Kreise umfassen, ergeben sich bei den Fallzahlen der Ausländer keine größeren prozentualen Abweichungen zur Hochrechnung auf Anpassungsschichtebene. Für einige Anpassungsschichten und Kreise sind die Ergebnisse der Hochrechnung in den Tabellen 2 und 3 veranschaulicht. Neben einigen größeren Kreisen, die gleichzeitig eine Anpassungsschicht bilden, wurde Bonn als kleinster Kreis, der ebenfalls eine Anpassungsschicht bildet, ausgewählt. Darüber hinaus wurde die Anpassungsschicht Coesfeld/Warendorf/Münster sowie Höxter/Paderborn gewählt. Beide Anpassungsschichten beinhalten Kreise mit besonders niedrigem und mittlerem Ausländeranteil. Es wird deutlich, dass sowohl bei den absoluten Zahlen (Tabelle 2) als auch bei den Anteilen (Tabelle 3) eine sehr hohe Übereinstimmung der Ergebnisse erreicht wird.

Abweichungen zu den Ergebnissen auf Anpassungsschichtebene ergeben sich immer dann, wenn bei der Bildung der Hochrechnungsfaktoren in bestimmten MZKR, die Teil dieser Anpassungsschicht sind, die 10-Fälle-Grenze zum Tragen kommt. Da dieses in NRW nicht der Fall ist, wurde das Problem am Beispiel von Bayern betrachtet. Für keine der Anpassungsschichten in Bayern ist für die Bildung der Hochrechnungsfaktoren auf Anpassungsschichtebene die 50-Fälle-Grenze zum Tragen gekommen (vgl. Tabelle 1). D.h. es wurden auf dieser Ebene keine Zusammenfassungen bei den ausländischen und deutschen Frauen bzw. Männern vorgenommen. Bei den MZKR hingegen wird in 7 MZKR die 10-Fälle-Grenze nicht erreicht. Entsprechend kommt es zwischen den hochgerechneten Ergebnissen auf Anpassungsschichtebene zu Abweichungen, wenn die Anpassungsschicht eine MZKR enthält, die keine gesonderten Hochrechnungsfaktoren für ausländische Männer und Frauen ausweist. Insgesamt finden sich in 5 der 22 Anpassungsschichten Abweichungen von über 10% in der Kategorie männliche und/oder weibliche Ausländer zwischen den hochgerechneten Ergebnissen mit dem Verfahren für MZKR und dem Verfahren auf Anpassungsschichtebene. Die höchste Abweichung in einer der betroffenen Anpassungsschichten beträgt ca. 36% in der Kategorie der männlichen Ausländer bezogen auf absolute Zahlen (vgl. Tabelle 2a im Anhang 1). Bezogen auf die Anteile liegen die Unterschiede der beiden Hochrechnungsverfahren in den 5 betroffenen Anpassungsschichten hingegen jeweils unter einem Prozentpunkt (vgl. Tabelle 3a im Anhang 1).

Für die neuen Bundesländer stellt sich dieser Sachverhalt anders dar. Obwohl im Vergleich zu Bayern in wesentlich mehr MZKR die Mindestbesetzung von 10 Fällen für die ausländischen Frauen bzw. Männer nicht erreicht wird, unterscheiden sich die für die Anpassungsschichten ermittelten Schätzwerte auf Basis der beiden Hochrechnungsverfahren kaum. Dies ist dadurch bedingt, dass auch bei dem

ANPS-Hochrechnungsverfahren in keinem der ostdeutschen Bundesländer gesonderte Hochrechnungsfaktoren für ausländische Männer und Frauen ausgewiesen werden, da in keiner der insgesamt 27 Anpassungsschichten (vgl. Tab. 1) die untere Grenze von 50 Fällen erreicht wurde. Entsprechend unterscheiden sich die Schätzergebnisse der beiden Hochrechnungsverfahren allenfalls geringfügig<sup>4</sup>, zugleich führen aber beide Verfahren zum Teil zu deutlichen Abweichungen von den Eckwerten für ausländische Frauen und Männer aus der laufenden Bevölkerungsfortschreibung.

---

<sup>4</sup> Nennenswerte Abweichungen treten nur dann auf, wenn eine Anpassungsschicht eine MZKR mit separaten Hochrechnungsfaktoren für ausländische Männer und Frauen enthält, für diese Anpassungsschicht selbst jedoch ein solcher separater Hochrechnungsfaktor nicht vorliegt, da die Mindestbesetzungszahl von 50 in der Anpassungskategorie nicht erreicht wurde.

Tabelle 2: Ausgewählte Anpassungsschichten/Kreise: Bevölkerung am Ort der Hauptwohnung (absolute Häufigkeiten)

AS-Nr.	KR-Nr.	Anpassungsschicht (AS) <sup>1</sup> Kreis (KR)	Mikrozensus ungewichtet				Mikrozensus gewichtet MZKR-Faktor				Mikrozensus gewichtet Anpassungsschichtfaktor			
			deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen
			abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.
<b>0503</b>		<b>Duisburg</b>	1.708	2.035	294	291	204.429	225.442	48.306	41.616	204.600	225.599	48.300	41.600
<b>0505</b>		<b>Essen</b>	2.222	2.540	191	207	250.259	281.007	36.091	32.158	250.400	281.200	36.000	32.100
<b>0508</b>		<b>Düsseldorf</b>	1.939	2.199	290	300	218.039	251.598	50.591	48.627	217.899	251.599	50.500	48.500
<b>0517</b>		<b>Bonn</b>	881	1.066	117	108	116.302	132.630	27.114	25.002	116.300	132.700	27.000	24.900
<b>0520</b>		<b>LK Coesfeld, LK Warendorf, LK Münster</b>	2.850	3.146	153	128	336.977	366.000	27.593	25.419	336.539	365.962	27.500	25.300
	05558	<i>LK Coesfeld</i>	842	877	14	14	99.036	104.220	4.997	4.527				
	05570	<i>LK Warendorf</i>	1.043	1.132	94	71	124.180	131.052	12.532	10.775				
	05515	<i>LK Münster</i>	965	1.137	45	43	113.761	130.728	10.064	10.117				
<b>0526</b>		<b>LK Höxter, LK Paderborn</b>	1.863	1.899	83	75	202.830	211.762	15.757	13.429	202.600	211.600	15.600	13.300
	05762	<i>LK Höxter</i>	715	702	12	18	72.320	75.382	4.316	3.650				
	05774	<i>LK Paderborn</i>	1.148	1.197	71	57	130.510	136.380	11.441	9.779				
		NRW gesamt	65.160	71.728	6.203	5.857	7.663.253	8.291.935	1.086.164	958.461	7.659.863	8.290.432	1.083.100	955.300

Quelle: Mikrozensus 2000

<sup>1</sup> Anpassungsschichten sind in der Tabelle „fett“, Kreise „kursiv“ gekennzeichnet. Anpassungsschichten, die gleichzeitig Kreise darstellen sind „fett“ und „kursiv“ gekennzeichnet.

Tabelle 3: Ausgewählte Anpassungsschichten/Kreise: Bevölkerung am Ort der Hauptwohnung ( in Prozent)

AS-Nr.	KR-Nr.	Anpassungsschicht (AS) <sup>1</sup> <i>Kreis (KR)</i>	Mikrozensus ungewichtet				Mikrozensus gewichtet MZKR-Faktor				Mikrozensus gewichtet Anpassungsschichtfaktor			
			deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen
			%	%	%	%	%	%	%	%	%	%	%	%
<b>0503</b>		<b>Duisburg</b>	39,5	47,0	6,8	6,7	39,3	43,4	9,3	8,0	39,3	43,4	9,3	8,0
<b>0505</b>		<b>Essen</b>	43,1	49,2	3,7	4,0	41,7	46,9	6,0	5,4	41,8	46,9	6,0	5,4
<b>0508</b>		<b>Düsseldorf</b>	41,0	46,5	6,1	6,3	38,3	44,2	8,9	8,5	38,3	44,3	8,9	8,5
<b>0517</b>		<b>Bonn</b>	40,6	49,1	5,4	5,0	38,6	44,1	9,0	8,3	38,7	44,1	9,0	8,3
<b>0520</b>		<b>LK Coesfeld, LK Warendorf, LK Münster</b>	45,4	50,1	2,4	2,0	44,6	48,4	3,6	3,4	44,6	48,5	3,6	3,3
	05558	<i>LK Coesfeld</i>	48,2	50,2	0,8	0,8	46,5	49,0	2,3	2,1				
	05570	<i>LK Warendorf</i>	44,6	48,4	4,0	3,0	44,6	47,0	4,5	3,9				
	05515	<i>LK Münster</i>	44,1	51,9	2,1	2,0	43,0	49,4	3,8	3,8				
<b>0526</b>		<b>LK Höxter, LK Paderborn</b>	47,5	48,4	2,1	1,9	45,7	47,7	3,6	3,0	45,7	47,8	3,5	3,0
	05762	<i>LK Höxter</i>	49,4	48,5	0,8	1,2	46,5	48,4	2,8	2,3				
	05774	<i>LK Paderborn</i>	46,4	48,4	2,9	2,3	45,3	47,3	4,0	3,4				
		NRW gesamt	43,7	48,2	4,2	3,9	42,6	46,1	6,0	5,3	42,6	46,1	6,0	5,3

Quelle: Mikrozensus 2000

<sup>1</sup> Anpassungsschichten sind in der Tabelle „fett“, Kreise „kursiv“ gekennzeichnet. Anpassungsschichten, die gleichzeitig Kreise darstellen sind „fett“ und „kursiv“ gekennzeichnet.

*Haushalte:*

Die Prüfung des MZKR-Haushaltshochrechnungsfaktors wurde bisher nur für NRW durchgeführt. Insgesamt führen beide Hochrechnungsverfahren zu vergleichbaren Schätzergebnissen. Für das gesamte Gebiet NRW erhält man mit den MZKR-Hochrechnungsfaktoren 7149 Haushalte mehr, als bei der Hochrechnung mit dem Anpassungsschichtfaktoren. Dies entspricht einer Abweichung von 0,01%. Aber auch die Zahl der Haushalte in den einzelnen Anpassungsschichten weist nur in wenigen Fällen Abweichungen von mehr als 0,1% auf. Geringfügig höhere Werte (1,1 bis 4,7%) treten nur bei 5 Anpassungsschichten auf, die sich aus zwei oder mehr Kreisen mit geringem Ausländeranteil zusammensetzen.

*Beruflicher Abschluss:*

Abschließend wird für den Vergleich der beiden Hochrechnungsverfahren das Merkmal ‚beruflicher Abschluss‘ betrachtet. Für dieses Merkmal erfolgt keine Anpassung im Rahmen des Hochrechnungsverfahrens, es steht aber in einem engen Zusammenhang mit der Staatsangehörigkeit einer Person. Das Merkmal „beruflicher Abschluss“ wurde auf Grundlage der Variablen „Beruflicher Ausbildungs- oder Hochschul-/Fachhochschulabschluss vorhanden?“ und „Höchster beruflicher Ausbildungs- oder Hochschul-/Fachhochschulabschluss“ gebildet und umfasst 3 Kategorien<sup>5</sup>, die im Folgenden mit den Begriffen „kein Abschluss“, „Ausbildung“ und „Hochschulabschluss“ charakterisiert werden.

In Tabelle 4 sind – bezogen auf NRW - die Anteilswerte von Deutschen und Ausländern innerhalb der einzelnen Ausbildungskategorien (% von höchster beruflicher Abschluss) sowie die prozentuale Verteilung der Abschlüsse innerhalb der deutschen und der ausländischen Population (% von Staatsangehörigkeit) dargestellt. Als wichtigstes Ergebnis lässt sich festhalten, dass der Einfluss der Hochrechnungsverfahren auf die Bildungsverteilung von Deutschen und Ausländern nahezu gegen Null geht. Beide Hochrechnungsverfahren führen zu weitgehend identischen Verteilungen, die sich wiederum kaum von den ungewichteten Ergebnissen unterscheiden. Ein etwas anderes Bild ergibt sich erwartungsgemäß in Hinblick auf die Verteilung ‚Deutsche-Ausländer‘ innerhalb der einzelnen Bildungsgruppen. Zwar sind die hochgerechneten Verteilungen wiederum nahezu identisch, aber in Hinblick auf die ungewichteten Ergebnisse kommt es zu Abweichungen. Im Vergleich zu den ungewichteten Schätzungen erhöht sich durch die Hochrechnung der Anteil der ausländischen Population in den einzelnen Bildungskategorien um 1,7 bis zu 5,3 Prozentpunkte.

---

<sup>5</sup> Kategorien: 1) Kein Abschluss 2) Ausbildung/Meister/Techniker: Zusammenfassung der Personen mit Anlernausbildung oder beruflichem Praktikum, Abschluss einer Lehrausbildung oder Vorbereitungsdienst für den mittleren Dienst in der öffentlichen Verwaltung, berufsqualifizierendem Abschluss an einer Berufsfachschule/Kollegschele oder Abschluss einer 1 jährigen Schule des Gesundheitswesens, Meiser-/Techniker oder gleichwertiger Abschluss einer 2 oder 3 jährigen Schule des Gesundheitswesens sowie Abschluss einer Fach- oder Berufsakademie und Abschluss einer Verwaltungshochschule sowie Personen mit einen Abschluss der Fachschule der ehemaligen DDR 3) Fach-/Hochschulabschluss/Promotion.

Tabelle 4 : Beruflicher Abschluss der Personen in Prozent – NRW-

Beruflicher Abschluss	Mikrozensus ungewichtet		Mikrozensus gewichtet MZKR-Faktor		Mikrozensus gewichtet Anpassungsschichtfaktor	
	deutsch	ausländ.	deutsch	ausländ.	deutsch	ausländ.
<b>Keine Abschluss:</b>						
Anzahl	29.839	5.215	3.473.309	877.728	3.472.227	874.141
% von höchster berufl. Abschluss	85,1	14,9	79,8	20,2	79,9	20,1
% von Staatsang.	28,8	63,4	28,7	62,7	28,7	62,8
<b>Ausbildung/Meister</b>						
<b>Techniker:</b>						
Anzahl	63.334	2.448	7.382.594	423.430	7.380.119	420.585
% von höchster berufl. Abschluss	96,3	3,7	94,6	5,4	94,6	5,4
% von Staatsang.	61,1	29,8	61,1	30,2	61,1	30,2
<b>FH-/Hochschulabschluss/Promotion:</b>						
Anzahl	10.464	560	1.227.065	99.762	1.226.546	97.580
% von höchster berufl. Abschluss	94,9	5,1	92,5	7,5	92,6	7,4
% von Staatsang.	10,1	6,8	10,2	7,1	10,2	7,0

Quelle: Mikrozensus 2000, Personen am Ort der Hauptwohnung.

Betrachtet man die Ergebnisse einzelner Anpassungsschichten so ergibt sich ein ähnliches Bild. In den meisten Anpassungsschichten liegen die Abweichungen zwischen den beiden Hochrechnungsverfahren bei weniger als einem Prozentpunkt. Etwas höhere Abweichungen finden sich für Anpassungsschichten, in welchen Kreise mit besonders niedrigem Anteil an ausländischen Personen zusammengefasst sind. Vergleichbar sind die Ergebnisse der beiden Hochrechnungsverfahren ebenfalls bei der Betrachtung der Anteile in Bezug auf den beruflichen Abschluss und das Geschlecht der Person.

### 2.3.3 Zusammenfassung der Ergebnisse

Insgesamt ist festzuhalten, dass mit den MZKR-Hochrechnungsfaktoren in Bezug auf die Variablen Staatsangehörigkeit und Geschlecht erwartungsgemäß eine gute Anpassung an die Zahlen der Bevölkerungsfortschreibung erreicht wird. Gleichfalls weisen die mit den verschiedenen Hochrechnungsverfahren ermittelten Schätzergebnisse eine hohe Konsistenz auf. So entsprechen die mit dem MZKR-Hochrechnungsverfahren auf Ebene der Anpassungsschichten ermittelten Schätzwerte nahezu

den Befunden des ANPS-Hochrechnungsverfahrens. Mit anderen Worten, beide Verfahren liefern vergleichbare Ergebnisse.

Abweichungen zwischen den beiden Hochrechnungsverfahren treten vorwiegend dann auf, wenn aufgrund der 10-Fälle Grenze keine getrennten Hochrechnungsfaktoren für ausländische und deutsche Personen ausgewiesen werden können. In den alten Bundesländern tritt dieses Problem nur bei einigen wenigen MZKR auf. In den neuen Bundesländern sind von diesem Problem mehrere Kreise betroffen. Jedoch ergeben sich die Abweichungen hier vor allem dadurch, dass für manche der MZKR in Ostdeutschland getrennte Hochrechnungsfaktoren ermittelt werden können, während bei der Hochrechnung auf Ebene der Anpassungsschichten – aufgrund der höheren Mindestbesetzungszahl (50-Fälle-Grenze) – keine getrennten Hochrechnungsfaktoren ausgewiesen werden. Insofern Konsistenz der verschiedenen Hochrechnungsverfahren angestrebt wird, erscheint es daher sinnvoll, für die neuen Bundesländer keine getrennten Hochrechnungsfaktoren für ausländische Frauen und Männer auf MZKR-Ebene auszuweisen.

Die Bereitstellung separater Hochrechnungsfaktoren auf MZKR-Ebene ermöglicht die Durchführung inhaltlicher Analysen für regionale Aggregationen, die nunmehr unabhängig von den bisherigen Aggregationen in den Anpassungsschichten vorgenommen werden können. Beim Arbeiten mit den Daten des Mikrozensus auf unterschiedlichen regionalen Aggregationsebenen ist jedoch darauf zu achten, welche Hochrechnungsfaktoren zur Anwendung kommen. Sofern man mit den Hochrechnungsfaktoren auf MZKR-Ebene arbeitet, muss man davon ausgehen, dass sich teilweise Unterschiede zu bereits veröffentlichten Ergebnissen auf Anpassungsschichtebene und zu den Zahlen der Bevölkerungsfortschreibung ergeben können. Dieser Sachverhalt ist allerdings grundsätzlich bei den Scientific-Use-Files gegeben und tritt auch bei den gegenwärtig verfügbaren Mikrozensus Scientific Use Files auf (Bihler, 2006).



### 3 Stichprobenfehler

Ein weiterer Aspekt, der hier angesprochen werden soll, ist der Stichprobenfehler. Mit einem Stichprobenumfang von einem Prozent der Personen und Haushalte ermöglicht der Mikrozensus die differenzierte Analyse auch kleinerer Bevölkerungsgruppen. Aufgrund seines hohen Auswahlsatzes sowie der Annahme, dass mit diesen Daten – auch wegen des geringen Unit-Nonresponse - Verteilungs- und Zusammenhangsmaße für die Grundgesamtheit verlässlich geschätzt werden können, wird der Mikrozensus oftmals als Referenzstatistik für wissenschaftsbasierte Bevölkerungserhebungen herangezogen. Dabei ist jedoch zu berücksichtigen, dass je kleiner der Stichprobenumfang ist, mit einem im Mittel umso größeren Stichprobenfehler bei Populations-/Punktschätzungen zu rechnen ist.<sup>6</sup> Dies trifft in besonderem Maße auf regionalisierte Analysen zu, bei welchen der Stichprobenumfang auf jene Untersuchungseinheiten reduziert ist, die der jeweiligen Regionaleinheit entstammen (Müller 1992:179). Von Seiten der amtlichen Statistik werden Ergebnisse mit einem relativen Standardfehler von über 15% aufgrund der geringen Aussagekraft nicht publiziert. Auf Landesebene werden beispielsweise Tabellenfelder mit Besetzungszahlen von unter 5.000 nicht ausgewiesen. Tabellenfelder mit Werten zwischen 5.000 und 10.000 werden wegen ihrer eingeschränkten Aussagekraft geklammert. Gleiches gilt für die auf dieser Basis errechneten Anteilswerte.

Bei kleinräumigen und gleichzeitig sachlich stark differenzierten Populationsschätzungen können die Fahlzahlen sehr schnell die oben angeführte Grenze unterschreiten. Auch wenn die Genauigkeitsanforderungen der amtlichen Statistik für die Forschung nicht bindend sind, ist es wichtig sich der Grenzen der regionalen Analyse in Hinblick auf die Aussagefähigkeit der Daten bewusst zu sein. Zu diesem Zweck wurden für alle 16 Bundesländer Fehlerrechnungen für ausgewählte Merkmale exemplarisch für jeweils zwei sehr große und zwei sehr kleine Mikrozensuskreisregionen im jeweiligen Bundesland durchgeführt.<sup>7</sup> Gewissermaßen als Daumenregel lässt sich festhalten, dass unabhängig von der Größe der betrachteten Regionaleinheit und den jeweiligen prozentualen Anteilen der einfache relative Standardfehler bei einer Besetzungszahl von unter 5.000 Fällen in der Population (50 Fälle in der Stichprobe) über 15 Prozent ansteigt. Beispielhaft und vereinfacht soll dieser Sachverhalt für unterschiedlich große Regionaleinheiten anhand von Tabelle 5 und 6 verdeutlicht werden.

Für die Berechnung des (relativen) Stichprobenfehlers ist zu berücksichtigen, dass es sich beim Mikrozensus nicht um eine uneingeschränkte Zufallsauswahl handelt, sondern um eine geschichtete Klumpenstichprobe. Zunächst wird der Standardfehler (s) unter Annahme einer uneingeschränkten

---

<sup>6</sup> Für die Qualitätsbeurteilung der Schätzung ist nicht nur die Größe des Stichprobenfehlers maßgeblich, sondern gleichfalls sind ‚non sampling errors‘ bzw. systematische Fehler von Bedeutung. Sogenannte ‚non sampling errors‘ entstehen bspw. durch Befragungsausfälle, Erinnerungsfehler, Untererfassung, Kodierfehler etc. Im vorliegenden Beitrag konzentrieren wir uns jedoch auf den Stichprobenfehler. Der Einfluss systematischer Fehlerquellen bleibt unberücksichtigt.

<sup>7</sup> Die entsprechenden Tabellen befinden sich in dem PDF-Dokument: Dokumentation zum Stichprobenfehler.

Zufallsauswahl (simple random sample: SI) und der Binomialverteilung der Absolutwerte (Totals) wie folgt ermittelt<sup>8</sup> :

$$s_{SI} = \sqrt{\frac{(1-f)}{(n-1)} * p_g (1-p_g)}$$

wobei

f = 0,01;

n = Anzahl der Stichprobenfälle insgesamt;

p<sub>g</sub> = n<sub>g</sub>/n: Anteil der Fälle der Merkmalskategorie g an allen Stichprobenfällen.

Der relative Standardfehler (RSF) ergibt sich aus dem Standardfehler des Erwartungswertes im Verhältnis zum Erwartungswert:

$$RSF_{SI} = s_{SI} / p_g$$

Die Schichtung und Klumpung im Mikrozensus werden bei der Berechnung des Stichprobenfehlers näherungsweise durch so genannte Design-Effekte (Zuschlagsfaktoren) berücksichtigt. Der Design-Effekt k(p) entspricht dem Verhältnis des design-basierten Standardfehlers (s) zum Standardfehler einer Stichprobe gleichen Umfangs unter Annahme einer uneingeschränkter Zufallsauswahl (s<sub>SI</sub>):

$$k_g = s(p_g) / s_{SI}(p_g)$$

Die amtliche Statistik schätzt den Design-Effekt für verschiedene Merkmalsgruppen mittels einer linearen Regression. Bis 1999 wurden Zuschlagsfaktoren für folgenden Gruppen ausgewiesen: (1) Bevölkerung und Erwerbstätige, ohne Ausländer oder Beschäftigte in Land- und Forstwirtschaft (2) Bevölkerung und Erwerbstätige sofern Ausländer oder Beschäftigte in Land- und Forstwirtschaft (3) Haushalte (vgl. Statistisches Bundesamt 1999:21).<sup>9</sup>

Der relative Standardfehler – mit näherungsweise Berücksichtigung des Stichprobendesigns – ergibt sich durch Multiplikation des geschätzten Design-Effekts bzw. Zuschlagsfaktor (k<sub>g</sub>) mit dem geschätzten relativen Standardfehler unter Annahme einer uneingeschränkten Zufallsauswahl (RSF):

$$RSF = RSF_{SI} * k_g$$

---

<sup>8</sup> Zu den folgenden Ausführungen vgl. u.a. Schimpl-Neimanns/Müller (2001) und Rendtel/Schimpl-Neimanns (2001).

<sup>9</sup> Ab 2000 wird die Fehlerrechnung nicht mehr ausschließlich für die drei genannten Gruppen durchgeführt. Für den Mikrozensus 2000 werden beispielsweise Fehlerkurven für Bevölkerung und Erwerbstätige, ohne Ausländer oder Beschäftigte in Land- und Forstwirtschaft und Bevölkerung und Erwerbstätige sofern Ausländer oder Beschäftigte in Land- und Forstwirtschaft unterschieden nach Ost- und Westdeutschland sowie für das gesamte Bundesgebiet zur Verfügung gestellt (vgl. Statistisches Bundesamt 2000:21).

In Tabelle 5 ist der relative Standardfehler für in der Größe variierende Regionaleinheiten und dem jeweiligen prozentualen Anteil einer Gruppe in der Regionaleinheit abgebildet. Wobei hier vereinfachend ein konstanter Zuschlagsfaktor für alle Anteilswerte unterstellt wird.<sup>10</sup> Tabelle 5 lässt sich entnehmen, dass der Standardfehler bei einem gegebenen Gruppenanteil mit zunehmender Größe der Regionaleinheit sinkt. In einer Regionaleinheit mit 100.000 Einwohnern ist bei einem Gruppenanteil von einem Prozent der relative Standardfehler näherungsweise 38 Prozent. Umfasst die Regionaleinheit hingegen 600.000 Einwohner liegt der entsprechende Wert nur bei knapp 16 Prozent. Ein relativer Standardfehler von ca. 11 Prozent wird bei einer Regionaleinheit von 600.000 schon bei einem Gruppenanteil von 2 Prozent erreicht, bei einer Regionaleinheit mit 100.000 Personen hingegen erst bei einem Gruppenanteil von 10 Prozent.

Tabelle 5: Relativer Standardfehler (RSF) in Abhängigkeit der Größe der Regionaleinheit und dem jeweiligen Anteil einer Gruppe  $p_g$  in der Regionaleinheit

Größe der Regional- einheit	$p_g$ in Regionaleinheit:							
	0,005	0,01	0,02	0,03	0,05	0,10	0,20	0,50
	in %							
	0,5	1	2	3	5	10	20	50
Relativer Standardfehler (RSF) in %								
600.000	22,1	15,6	11,0	8,9	6,8	4,7	3,1	1,6
500.000	24,2	17,1	12,0	9,8	7,5	5,2	3,4	1,7
400.000	27,1	19,1	13,4	10,9	8,4	5,8	3,8	1,9
300.000	31,3	22,1	15,5	12,6	9,7	6,6	4,4	2,2
200.000	38,3	27,0	19,0	15,4	11,8	8,1	5,4	2,7
100.000	54,2	38,2	26,9	21,8	16,7	11,5	7,7	3,8

RSF =  $RSF_{SI} \cdot k_g$ ; Vereinfachend wird hier  $k_g = 1,22$  eingesetzt

Anschaulich im Hinblick auf die Entwicklung des Stichprobenfehlers bei unterschiedlich großen Regionaleinheiten ist die Betrachtung des relativen Standardfehlers bei variierender Größe der Regionaleinheit in Abhängigkeit der Besetzungszahl einer Gruppe. Wie aus Tabelle 6 ersichtlich, ist der relative Standardfehler bei einer gegebenen Gruppengröße nahezu unabhängig von der Größe der Regionaleinheit. Bei einer Gruppengröße von 300 Fällen (diese entsprechen im Datensatz ca. 3 Fällen) liegt der relative Standardfehler in allen Regionaleinheiten über 70%. Bei einer Besetzungszahl von 1000 Fällen liegt der Wert noch immer bei circa 38 Prozent. Erst ab einer Gruppengröße von 5.000 Fällen (ca. 50 Fällen in der Stichprobe) wird ein Wert von unter 20 Prozent für alle Regionaleinheiten erreicht. Das entspricht in einer Regionaleinheit von 600.000 einem Gruppenanteil von circa 0,8 Prozent und in einer Regionaleinheit von 100.000 einem Anteil von 5 Prozent. Entscheidend bei dieser Betrachtungsweise ist jedoch die Besetzungszahl in einer Teilgruppe ( $n_g$ ), denn diese verdeutlicht nochmals, dass bei einer Zellbesetzung unter 50 Fällen im Mikrozensus, die Präzision von Populations-schätzungen erheblich eingeschränkt ist und zwar unabhängig davon ob die Schätzungen im Bundes-

<sup>10</sup> De facto ist der Zuschlagsfaktor keine Konstante, sondern steigt mit dem relativen Anteil  $p_g$  annähernd linear an (vgl. Statistisches Bundesamt 1996).

durchschnitt, auf Ebene der Anpassungsschichten oder der Mikrozensuskreisregionen vorgenommen werden.

Tabelle 6: Relativer Standardfehler (RSF) in Abhängigkeit der Größe der Regionaleinheit bei gegebener Besetzungszahl einer Gruppe (n) in Regionaleinheit

Größe Regional- Einheit	(frei hochgerechnete) Besetzungszahl (n <sub>g</sub> ) in Regionaleinheit:							
	300	500	1000	5000	10.000	20.000	50.000	100.000
	Relativer Standardfehler in %							
600.000	70,1	54,3	38,4	17,1	12,0	8,4	5,2	3,5
500.000	70,1	54,3	38,4	17,1	12,0	8,4	5,2	3,4
400.000	70,1	54,3	38,3	17,1	12,0	8,4	5,1	3,3
300.000	70,1	54,3	38,3	17,0	11,9	8,3	5,0	3,1
200.000	70,0	54,2	38,3	17,0	11,8	8,1	4,7	2,7
100.000	70,0	54,2	38,2	16,7	11,5	7,7	3,8	-

RSF = RSF<sub>SI</sub> \* k<sub>g</sub>; Vereinfachend wird hier k<sub>g</sub> = 1,22 eingesetzt

In Hinblick auf die Präzision von Punktschätzungen, erweist sich die Faustregel die Besetzungszahl von 50 Fällen nicht zu unterschreiten daher sowohl für Analysen auf Basis des Mikrozensus-Regionalfiles wie auch des Mikrozensus-Grundfiles, als sinnvoll. Zur Problemverdeutlichung sei nochmals auf die Dokumentation zum Stichprobenfehler hingewiesen, in welcher für ausgewählte Merkmale das Ausmaß des relativen Stichprobenfehlers sowie die Konfidenzintervalle länderspezifisch jeweils für zwei sehr große und zwei sehr kleine Mikrozensuskreisregionen beispielhaft dokumentiert ist.

## 4 Anonymisierungsmaßnahmen

### 4.1 Die Anonymisierungsempfehlungen

Die Anonymisierungsmaßnahmen für das Mikrozensus-Regionalfile orientieren sich an den Vorgaben des Projektes ‚Faktische Anonymisierung von Mikrodaten‘ (Müller et al., 1991), und zielen maßgeblich auf eine Reduzierung der sachlichen Differenzierung bei einer tiefen regionalen Differenzierung ab (vgl. Übersicht 1). Auf eine Begründung der jeweiligen Anonymisierungsmaßnahmen wird an dieser Stelle verzichtet, da sie bei Müller et al. (1991) ausführlich dokumentiert ist. Die Operationalisierung der Anonymisierungsempfehlungen orientiert sich an der Prämisse unter Einhaltung der Datenschutzbestimmungen, möglichst wenig in das Analysepotenzial der Daten einzugreifen. Im Wesentlichen wird hierbei angestrebt, die potenziellen Forschungsinteressen möglichst breit abzudecken, d.h. die Anonymisierungsmaßnahmen so zu gestalten, dass nach wie vor eine flexible Nutzung der Daten für die Bearbeitung der jeweils interessierenden Fragestellung möglich ist. In wissenschaftlichen Analysen steht oftmals die statistische Modellierung, die durch möglichst differenzierte Angaben im Regionalfile ermöglicht wird, im Vordergrund. So sind etwa Arbeitsmarkt- und Ungleichheitsforscher häufig an einer sehr differenzierten Ausweisung z.B. der Berufsangabe interessiert, um sie in Kombination mit anderen Informationen zur Erwerbstätigkeit für die Operationalisierung spezifischer theoretischer Konstrukte (z.B. Klassenlage, Schicht, sozio-ökonomischer Status, Berufsprestige) zu nutzen, welche dann als Prädiktor- oder Kriteriumsvariablen in die eigentlichen Analysen Eingang finden. Vor diesem Hintergrund steht bei der Umsetzung der Anonymisierungsmaßnahmen auf kleinräumiger Ebene, nicht die Aussagekraft von Punktschätzungen im Vordergrund, sondern die Flexibilität des Analysepotenzials.

Übersicht 1: Datenfilespezifische Anonymisierungsempfehlungen: Mikrozensus-Regionalfile

<b>Datenfilespezifische Anonymisierungsmaßnahmen: MIKROZENSUS-REGIONALFILE</b>	
1. <b>Tiefe regionale Differenzierung</b> , wobei gilt:	<ul style="list-style-type: none"> <li>• Es soll keine Regionaleinheit ausgewiesen werden, die weniger als 100.000 Einwohner aufweist</li> </ul>
2. <b>Merkmale Nationalität, Beruf, Wirtschaftszweig und Alter:</b>	<ul style="list-style-type: none"> <li>• Keine Ausweisung von Merkmalsausprägungen mit weniger als 50.000 Einwohnern in der Bundesrepublik</li> <li>• pro übermittelter Regionaleinheit muss jede Merkmalsausprägung mindestens 300 Fälle in der Population aufweisen</li> </ul>
3. <b>Sonstige Merkmale:</b>	<ul style="list-style-type: none"> <li>• Merkmalsvergrößerung, so dass jede ausgewiesene Merkmalsausprägung für die Bundesrepublik mindestens 5.000 Fälle umfasst</li> </ul>
4. <b>Stichprobenziehung:</b>	Nur eine Substichprobe von 85% wird weitergegeben

Quelle: Müller et al. 1991: XX.

Wie aus Übersicht 1 ersichtlich beschränken sich die Anonymisierungsempfehlungen hinsichtlich der Regionalisierung auf die Vorgabe eines Mindestbesetzungskriteriums pro kleinster ausgewiesener

Regionaleinheit: Es darf keine Regionaleinheit mit weniger als 100.000 Einwohnern ausgewiesen werden. Wie eingangs ausgeführt, wurde dieses Konzept durch die Generierung der so genannten Mikrozensuskreisregionen realisiert. Als Ausgangspunkt der MZKR dienen Kreise bzw. Kreisregionen. Diese wurden dahingehend überprüft, ob sie das durch die Anonymisierungsempfehlungen vorgegebene Kriterium von mindestens 100.000 Einwohnern pro kleinster ausgewiesener Regionaleinheit erfüllen. Ist das Mindestbesetzungskriterium erfüllt, werden die Kreise/Kreisregionen als Regionaleinheit übernommen. Bei Nichterfüllung des Mindestbesetzungskriteriums, werden die betroffenen Kreise/Kreisregionen innerhalb der übergeordneten Regionaleinheit ‚Raumordnungsregion‘ aggregiert. Innerhalb von Raumordnungsregionen werden wiederum Kreise/Kreisregionen des gleichen Kreistyps zusammengefasst. Insgesamt werden 349 MZKR ausgewiesen, die von den Forschern je nach den spezifischen Forschungsinteressen weiter aggregiert werden können. So ist bspw. ein Umstieg von MZKR zu Arbeitsmarktregionen durch Zusammenfassung möglich, bzw. einzelne MZKRs bilden ein Aggregat von Arbeitsmarktregionen. Insgesamt stellen die Mikrozensuskreisregionen für unterschiedlichste Analyseziele der Regionalforschung eine sinnvolle Untergliederung dar. Die Ausrichtung an Kreisen bzw. Kreisregionen gewährleistet zugleich eine hohe Kontinuität der MZKR im Zeitverlauf, als eine wichtige Voraussetzung für die Bearbeitung intertemporaler Fragestellungen. Darüber hinaus verringert eine regionale Abgrenzung, die nicht jährlich neu überdacht werden muss, den Zeit- und Kostenaufwand der Datenbereitstellung erheblich. Da eine Vielzahl von regionalen Indikatoren auf Kreisebene vorliegt, besteht weiterhin die Möglichkeit Makro- und Mikrodaten zu verknüpfen.

Generell wird nur eine Substichprobe von 85 Prozent des Originalmaterials weitergegeben.

Weiterhin gilt grundsätzlich, dass für alle Merkmale ein Besetzungsminimum von mindestens 5.000 Fällen in der Grundgesamtheit, d.h. 50 Fällen in der Stichprobe erfüllt sein muss. Diese Vorgabe gilt auch für das Mikrozensus-Grundfile.

Für die stark differenziert erfassten Merkmale ‚Beruf‘, ‚Wirtschaftszweig‘, ‚Nationalität‘ und ‚Alter‘ bzw. ‚Geburtsjahr‘ gilt darüber hinaus die Vorgabe, dass im Bundesdurchschnitt keine Ausprägung mit einer Besetzung unter 50.000 Einwohnern in der Grundgesamtheit (in der Stichprobe 500 Fälle) ausgewiesen werden darf. Bezogen auf die kleinste übermittelte Regionaleinheit, d.h. den MZKR, muss jede Ausprägung mindestens 300 Fälle in der Grundgesamtheit bzw. 3 Fälle in der Stichprobe aufweisen. Diese Vorgaben erweisen sich als vergleichsweise diffizil, da sowohl auf der Bundesebene wie auf der MZKR-Ebene umfangreiche Zusammenfassungen notwendig werden, um die jeweiligen Mindestbesetzungskriterien zu erfüllen.

Beispielhaft sei dies in Tabelle 7 für den Bundesdurchschnitt verdeutlicht. Von den 369 Berufsordnungen (3-Steller) erfüllen rund 60 Prozent das Mindestbesetzungskriterium von 500 Fällen nicht, beim Wirtschaftszweig trifft dies auf knapp die Hälfte der Ausprägungen zu, bei der Nationalität sind etwa 37 Prozent der Kategorien betroffen. Bezogen auf die Individuen zeichnet sich ein erheblich weniger dramatisches Bild ab. De facto repräsentieren die 223 kritischen Berufskategorien knapp 14 Prozent der erwerbstätigen Bevölkerung, beim Wirtschaftszweig und der Nationalität entsprechen die kritischen Kategorien gut 6 Prozent der Bevölkerung. Diese Diskrepanz zwischen dem hohen Ausmaß der kritischen Kategorien einerseits und des relativ geringen Anteils der betroffenen Population andererseits, ebenso wie das Bestreben das Analysepotenzial der Daten nicht unnötig einzuschränken, ist

ein wichtiger Grund dafür, keine pauschalen Zusammenfassungen der betroffenen Ausprägungen vorzunehmen, sondern eine problemorientierte Herangehensweise zu wählen.

Tabelle 7: Beruf, Wirtschaftszweig, Nationalität und Geburtsjahr: Anteil und Anzahl der Fälle, die das Mindestbesetzungskriterium (MBK) auf Bundesebene nicht erfüllen nach Merkmalskategorien und Individuen.

Variable	Untersuchungseinheit					
	Kategorien			Individuen		
	Total	MBK <sup>*)</sup> nicht erfüllt		Total	MBK <sup>*)</sup> nicht erfüllt	
		N	%		N	%
Berufsordnungen	369	223	60,4	321.004	44.668	13,9
Wirtschaftszweig-	222	113	50,9	321.004	20.328	6,3
Nationalität	38	14	36,8	53.918	3.493	6,5
Geburtsjahr	96	2	2,1	718.859	794	0,1

<sup>\*)</sup>MBK = Mindestbesetzungskriterium (50.000 Fälle in Population, bzw. 500 in Mikrozensus)

Quelle: Mikrozensus 2000; eigene Berechnungen

In Hinblick auf das Analysepotenzial des Regionalfiles wäre es dabei wünschenswert, merkmalspezifisch für jede einzelne der kritischen Kategorien anhand inhaltlicher Kriterien über eine Zusammenfassung mit anderen Kategorien zu entscheiden. Bei einer solchen Vorgehensweise ist jedoch zu berücksichtigen, dass die auf Bundesebene erfolgten Aggregationen mit den Zusammenfassungen auf der MZKR-Ebene kompatibel sein müssen. Zugleich sollten die Vergrößerungen auch für verschiedene Erhebungszeitpunkte weitgehend konstant gehalten werden, um intertemporal vergleichende Analysen zu ermöglichen. Beide Bedingungen sind bei Einzelfallentscheidungen – sofern überhaupt – nur mit einem immensen Arbeitsaufwand zu erreichen. Weniger zeitaufwändig, fehleranfällig und mit geringeren regions- und zeitspezifischen Vergleichbarkeitsproblemen behaftet, sind Routinen, bei welchen die kritischen Kategorien nach fest vorgegebenen Regeln zusammengefasst werden.

## 4.2 Beruf und Wirtschaftszweig

Unter dieser Maßgabe erweist es sich als Vorteil, dass die beiden Merkmale ‚Beruf‘ und ‚Wirtschaftszweig‘ einer hierarchischen Gliederung folgen, die für eine selektive Zusammenfassung von kritischen Kategorien genutzt werden kann (siehe ausführlich Wirth/Schunck 2007).

Hierfür werden in einem ersten Schritt für die vier Hierarchieebenen des Berufsmerkmals (Berufsordnungen, Berufsgruppen, Berufsabschnitte und Berufsbereiche)<sup>11</sup> jeweils eine Variable generiert. In

<sup>11</sup> Die Berufsordnungen sind durch eine dreistellige Kennziffern mit insgesamt 369 Kategorien charakterisiert. Berufsgruppen weisen eine zweistellige Kennziffer mit 88 Ausprägungen auf, die Berufsabschnitte umfassen 33 Ausprägungen, die Berufsbereiche weisen 6 Kategorien auf. Darüber hinaus

einem zweiten Schritt wird dann auf Bundesebene mittels einer Routine überprüft, welche Berufsordnungen das Mindestbesetzungskriterium nicht erfüllen. Diese kritischen Kategorien werden der Ausprägung ‚nicht näher ausgewiesen‘ zugeordnet. Sie können jedoch weiterhin in Form von Berufsgruppen, -abschnitten oder -bereichen für Analysen genutzt werden. Allerdings erweist sich diese Maßnahme als noch nicht hinreichend, da weiterhin in einigen Berufsgruppen und vereinzelt auch in Berufsabschnitten und Berufsbereichen das Mindestbesetzungskriterium nicht erfüllt ist. Für diese kritischen Kategorien wurde vorab eine Zusammenfassung auf Ebene der Berufsordnungen vorgenommen.

Analog zur Vorgehensweise auf Bundesebene wurden die vier hierarchischen Berufsmerkmale anschließend sukzessive in allen MZKR auf Erfüllung des Mindestbesetzungskriteriums überprüft. Regionenspezifisch werden zunächst jene Berufsordnungen, die unterhalb der kritischen Schwelle liegen, der Kategorie ‚nicht näher ausgewiesen‘ zugeordnet. Das gleiche Verfahren kommt anschließend bei den Berufsgruppen und den Berufsabschnitten zur Anwendung. Im Ergebnis bewirkt diese Herangehensweise einerseits, dass die Anonymisierung für alle MZKR der gleichen Systematik folgt, d.h. über die Regionaleinheiten hinweg konsistent und damit vergleichbar ist. Andererseits wird zugleich den regionenspezifischen Berufsstrukturen Rechnung getragen. Mit anderen Worten, weisen spezifische Berufsordnungen (z.B. Landwirte; Seeleute etc.) in einer oder mehreren Regionaleinheiten eine zu geringe Zellbesetzung auf, werden sie zwar in den betroffenen Regionen aggregiert, in Regionen mit ausreichend hoher Besetzung bleiben diese Berufsordnungen hingegen erhalten. Auf diese Weise wird sowohl dem Bedürfnis nach regionenspezifischen Analysen Rechnung getragen, als auch die interregionale Vergleichbarkeit des Merkmals Beruf gewährleistet.

Betrachtet man den Anteil der nach Anonymisierung auf Mikrozensuskreisebene noch verfügbaren Berufsordnungen (3-Steller) als Indikator für das verbleibende Analysepotenzial der Daten, kommt man zu einer eher negativen Einschätzung. Je nach MZKR werden von den ursprünglichen Berufsordnungen nur noch zwischen knapp 30 und gut 50 Prozent ausgewiesen. Da die verbleibenden Berufsordnungen jedoch 60 bis 90 Prozent der erwerbstätigen Bevölkerung repräsentieren, enthalten die Daten für die berufsspezifische Regionalforschung nach wie vor ein erhebliches Analysepotenzial. Ist das Forschungsinteresse auf Berufsgruppen (2-Steller) gerichtet, ergibt sich ein deutlich positiveres Bild. Der Anteil der nach Anonymisierung noch ausgewiesenen Berufsgruppen variiert je nach MZKR zwischen 60 und 90 Prozent, die korrespondierenden Anteile an den Erwerbstätigen liegen zwischen etwa 60 und fast 100 Prozent.

Anonymisierung geht immer mit einer Reduzierung des Informationsgehalts der Daten einher. In Bezug auf das Merkmal Beruf wurde versucht, diesen Informationsverlust so gering wie möglich zu halten. Die vorgenommenen Aggregationen orientieren sich an der amtlichen Berufssystematik, d.h. der Klassifikation der Berufe. Mit dieser Vorgehensweise werden vor allem jene Forschungsinteressen abgedeckt, die primär auf die Analyse von Berufen oder Berufstypisierungen ausgerichtet sind. Eine wesentliche Einschränkung ergibt sich jedoch für solche Forschungsinteressen, bei welchen die Be-

---

gibt es auch noch Berufsklassen mit einer vierstelligen Kennziffer. Diese werden jedoch in Scientific-Use-Files grundsätzlich nicht weitergegeben und daher im Folgenden nicht berücksichtigt.



rufangaben für die Operationalisierung eines theoretischen Konstrukts wie z.B. des sozio-ökonomischen Status oder der EGP-Klassifikation benötigt werden. Zwar wird bei der Generierung solcher Konstrukte gleichfalls eine Zusammenfassung von Berufen vorgenommen, diese orientiert sich jedoch in aller Regel nicht an der amtlichen Berufssystematik, sondern an inhaltlichen Kriterien, wie z.B. dem mit einem Beruf verbundenen Grad der Handlungsautonomie oder spezifischen Humankapital (vgl. Brauns et al. 2000; Hausen et al. 2006; Müller et al. 2007; Wirth et al. 2009). Für den Mikrozensus stehen mit der Erikson-Goldthorpe-Portocarero(EGP)-Klassifikation, der Berufsklassifikation von Blossfeld und der Magnitude-Prestige-Skala gegenwärtig drei solcher sozialwissenschaftlicher Konstrukte zur Verfügung, die auf Basis der KldB-Berufsordnungen operationalisiert wurden. Nach Anonymisierung der Berufsangaben im Mikrozensus-Regionalfiler lässt sich keines dieser Konstrukte mehr operationalisieren. Deshalb werden die für Operationalisierung der EGP-Klassifikation wie auch der Blossfeld-Berufsklassifikation notwendige Gruppierungen von Berufen vor der Anonymisierung der Berufsangaben erstellt und im Scientific-Use-File bereitgestellt. Auf dieser Basis kann die EGP-Klassifikation wie auch die Blossfeld Berufsklassifikation von den Forscher/innen selbst generiert und für inhaltliche Analysen genutzt werden. Da es sich hierbei um stark aggregierte Berufsgruppen handelt, die keinen Rückschluss auf den genauen Beruf der einzelnen Person ermöglichen, wird die Variable nach dem Vorbild der in Übersicht 1 dargestellten Maßnahmen für „sonstige Merkmale“ anonymisiert, d.h. keine der Merkmalsausprägungen der Variable weist auf Bundesebene weniger als 5.000 Fälle auf. In Hinblick auf das Forschungspotenzial der Daten gewinnt das Mikrozensus-Regionalfiler hierdurch erheblich<sup>12</sup>.

Analog zum Berufsmerkmal werden auch beim Wirtschaftszweig die Hierarchieebenen genutzt, um bei Einhaltung der Anonymisierungskriterien möglichst wenig in das Analysepotenzial der Daten einzugreifen. Als Ausgangsbasis für die Anonymisierung dienen die Wirtschaftsgruppen, aus welchen die Abteilungen, Unterabschnitte und Abschnitte generiert werden. Auch in diesem Fall sind wiederum vorab Zusammenfassungen der Wirtschaftsgruppen notwendig, da schon auf Bundesebene bei 14 von 60 Abteilungen, fünf von 31 Unterabschnitten und zwei von 17 Abschnitten das Mindestbesetzungskriterium nicht erfüllen. Anschließend wird die Erfüllung des Mindestbesetzungskriteriums auf Bundesebene zunächst für die Wirtschaftsgruppen, dann für die Abteilungen, Unterabschnitte und Abschnitte überprüft. Kategorien, deren Besetzungszahl unterhalb des Minimums liegen, werden – entsprechend der Vorgehensweise beim Berufsmerkmal – nicht ausgewiesen, sondern stehen jeweils nur noch in einer stärker aggregierten Form zur Verfügung. Als Ergebnis der Vorab-Zusammenfassung und der anschließenden sukzessiven Anonymisierungsroutine verringert sich die Anzahl der Wirtschaftsgruppen von 222 auf 113, der Abteilungen von 60 auf 47, der Unterabschnitte von 31 auf 27 und der Abschnitte von 17 auf 15 Kategorien. Von dieser Basis ausgehend werden die Zellbesetzungen der vier Wirtschaftszweigmerkmale im nächsten Schritt für alle Mikrozensuskreisregionen überprüft und sofern notwendig aggregiert. Nach Abschluss der Anonymisierung zeigt sich eine deutliche Variation zwischen den MZKR. Im Maximum bleiben weit über 60 Prozent der Katego-

---

<sup>12</sup> Eine entsprechende Bereitstellung von Prestigeskalen ist in Hinblick auf die Erhaltung des Analysepotenzials nicht weiterführend, da die Berufsordnungen gewissermaßen eins-zu-eins Prestigescores zugeordnet werden und dementsprechend ebenso umfangreiche Aggregationen notwendig sind, wie bei den ursprünglichen Berufsangaben.

rien, im Minimum etwas unter 40 Prozent der Kategorien der Wirtschaftsgruppen erhalten, welche respektive über 95 Prozent bzw. noch immerhin 75 Prozent der erwerbstätigen Bevölkerung abbilden (vgl. Abbildung 1 in Anhang 2). Bezogen auf die Wirtschafts-Abteilungen verbleiben je nach Region zwischen gut 90 Prozent und knapp 60 Prozent der Kategorien, wobei der Anteil der hierdurch repräsentierten Erwerbstätigen in allen Regionen deutlich über 90 Prozent liegt (vgl. Abbildung 2 im Anhang 2).

Der Beruf und Wirtschaftszweig der zweiten Erwerbstätigkeit wird nur auf 1-Stellerebene zur Verfügung gestellt. Wenn eine Merkmalsausprägung auf Bundesebene weniger als 5.000 Fälle umfasst, werden der Beruf und Wirtschaftszweig, sofern keine sinnvolle Zusammenfassung möglich ist, nicht näher ausgewiesen. Da in diesem Fall keine weiteren Informationen zum Beruf und Wirtschaftszweig der zweiten Erwerbstätigkeit im Datensatz enthalten ist und somit keine Rückschlüsse auf den konkreten Beruf und bzw. differenzierte Angaben zum Wirtschaftszweig möglich sind, kommen in diesem Fall die strengeren Anonymisierungskriterien für Beruf und Wirtschaftszweig nicht zum Tragen.

### 4.3 Alter

In Hinblick auf die Anonymisierung als relativ einfach handhabbar, erweisen sich die Merkmale ‚Alter‘ bzw. ‚Geburtsjahr‘, wobei wir uns hier exemplarisch auf die Altersvariable konzentrieren. Von den insgesamt 96 Ausprägungen (inkl. der nach oben offenen Randklasse ‚95 Jahre und älter‘) erfüllen auf Bundesebene nur die zwei Altersjahre ‚94‘ und ‚95‘ das Mindestbesetzungskriterium nicht. Ein etwas anderes Bild ergibt sich auf der Ebene der Mikrozensuskreisregionen. Während für die Altersjahre 0 bis 80 Jahre, das Mindestbesetzungskriterium in nahezu 100 Prozent der Regionaleinheiten erfüllt ist, geht dieser Anteil bei den über 80-Jährigen auf unter 90 Prozent zurück und liegt bei den 90-Jährigen nur bei etwa 50 Prozent. Um den anonymisierungsbedingten Informationsverlust wiederum möglichst gering zu halten, da auf MZKR-Ebene gelegentlich – wenn zwar auch äußerst selten – auch andere Alterskategorien unterhalb des Mindestbesetzungskriteriums liegen, wird daher zusätzlich zu der differenzierten Altersvariablen eine gruppierte Altersvariable mit 20 Kategorien (vgl. Übersicht 2) generiert. Im Weiteren werden in einem ersten Schritt auf Bundesebene, in einem zweiten Schritt auf der Ebene der MZKR alle Kategorien der detaillierten Altersvariablen, die das Mindestbesetzungskriterium nicht erfüllen, auf die Kategorie ‚nicht näher ausgewiesen‘ gesetzt und nur noch über die gruppierte Altersvariable ausgewiesen.

Übersicht 2: Ergänzende gruppierte Altersvariable im Mikrozensus-Regionalfile

Kategorie	Altersklasse
1	0-2
2	3-5
3	6-9
4	10-14
5	15-17
6	18-19
7	20-24
8	25-29
9	30-34
10	35-39
11	40-44
12	45-49
13	50-54
14	55-59
15	60-64
16	65-69
17	70-74
18	75-79
19	80-84
20	85 und älter

### 4.4 Nationalität

Da für das Merkmal ‚Nationalität‘ auch schon im Mikrozensus-Grundfile keine Nationalität oder Gruppen von Nationalitäten mit weniger als 50.000 Einwohnern im Bundesdurchschnitt (bzw. 500 Fällen im Mikrozensus) ausgewiesen werden, können die im Grundfile vorgenommenen Zusammenfassungen direkt für das Regionalfile übernommen werden (vgl. Übersicht 3).

Übersicht 3: Zusammenfassung der Nationalitäten für das Mikrozensus-Grundfile

<b>Staatsangehörigkeit</b>	<b>Kategorie</b>
Bosnien-Herzegowina	02
Frankreich	05
Griechenland	06
Großbritannien	07
GUS	08
Italien	10
Kroatien	11
Niederlande	13
Österreich	14
Polen	15
Portugal	16
Rumänien	17
Serbien und Montenegro	20
Slowakei, Tschechische Republik, Ungarn	21, 24
Spanien	22
Türkei	23
Übrige EU (Belgien, Dänemark, Finnland, Irland, Luxemburg, Schweden)	01, 03, 04, 09, 12, 18
Übriges Europa (Schweiz, Sonstiges Ost- und Mitteleuropa, Sonstiges Westeuropa)	19, 25, 26
Marokko	27
Sonstiges Afrika	28
Vereinigte Staaten von Amerika (USA)	29
Sonstiges Amerika	30, 31
Iran	32
Sonstiger Naher Osten (z.B. Libanon, Jordanien, Syrien, Irak, Israel)	33
Vietnam	34
Sonstiges Südasien (z.B. Afghanistan, Indien, Kambodscha, Laos, Pakistan, Thailand, Sri Lanka)	35
Ostasien (z.B. China, Hongkong, Indonesien, Japan, Korea, Macao, Philippinen)	36
Übrige Welt, Staatenlos	45, 50
Deutscher ohne weitere Staatsangehörigkeit	leer
Quelle: Schlüsselverzeichnis, Mikrozensus 2000; Scientific-Use-File	

Nach diesem Anonymisierungsschritt auf Bundesebene verbleiben auf Mikrozensuskreisebene je nach MZKR zwischen ein bis 12 Nationalitäten bzw. Gruppen von Nationalitäten bei welchen das Mindestbesetzungskriterium von 300 Fällen in der Grundgesamtheit (bzw. 3 Fällen in der Stichprobe) nicht erfüllt ist. Eine weitere Teilzusammenfassung wie dies für Beruf, Wirtschaftszweig und Alter geschehen ist, erweist sich für das Merkmal Nationalität als wenig hilfreich. Denn selbst bei einer Zusammenfassung nach Kontinenten, verbleiben in einzelnen MZKR noch immer kritische Kategorien. Hinzu kommt, dass eine solch grobe Zusammenfassung inhaltlich kaum noch Sinn macht. Deshalb werden auf Ebene der MZKR alle Nationalitäten, die das Mindestbesetzungskriterium nicht erfüllen, der Kategorie ‚Ausländer, nicht näher ausgewiesen‘ zugeordnet.

Zugleich ist festzuhalten, dass selbst bei einer nur dichotomen Differenzierung zwischen Deutschen und Ausländern die Fallzahlen die ausländische Population vor allem für Ostdeutschland sehr klein

werden. Populationsschätzungen sind auf dieser Basis für die ostdeutschen und manche der westdeutschen Mikrozensuskreisregionen nicht mehr aussagefähig. Dennoch ist es unter Berücksichtigung dieser Restriktionen, wichtig das Nationalitätenmerkmal auch auf Ebene der MZKR möglichst differenziert verfügbar zu machen, um das Analysepotenzial der Daten und den Anwendungsbereich in der Forschung nicht unnötig einzuschränken. Denn zum einen ist die Nationalität ein wichtiger Einflussfaktor für Bildungs-, Arbeitsmarkt- und Berufschancen und fließt als solcher in die statistische Modellierung ein. Zum anderen verbleibt der Forschung die Möglichkeit auf einer den jeweiligen Forschungsfragen entsprechenden höheren regionalen Aggregationsebene das Merkmal Nationalität zu nutzen. Und schließlich ergibt sich in längerfristiger Perspektive auch die Möglichkeit Daten unterschiedlicher Erhebungsjahrgänge zu kumulieren und damit die Fallzahlen zu erhöhen.

## 5 Fazit

Im Rahmen der Arbeiten am Mikrozensus-Regionalfile wurde ein leistungsfähiges Hochrechnungsverfahren auf Ebene der MZKR entwickelt, das sich stark an dem Hochrechnungsverfahren bis zum Mikrozensus 2004 orientiert und vergleichbare Ergebnisse produziert. Ziel war insbesondere, dass bei der Anwendung der unterschiedlichen Hochrechnungsverfahren auf Anpassungsschichtsebene und MZKR-Ebene konsistente Ergebnisse produziert werden. Das Programm liegt als SPSS-Routine vor und kann flexibel für alle Mikrozensusjahrgänge bis einschließlich 2004 umgesetzt werden.

Die regional differenzierten Hochrechnungsverfahren ermöglichen Populationsschätzungen für räumliche Aggregationen unabhängig von der Zusammensetzung der regionalen Anpassungsschichten. Durch diese Hochrechnungsfaktoren wird eine wesentlich flexiblere Auswertung des Mikrozensus ermöglicht. Es liegen beispielsweise für größere Kreise, die bisher mit anderen Kreisen in Anpassungsschichten zusammengefasst wurden, separate Hochrechnungsfaktoren vor. Es besteht weiterhin die Möglichkeit regionale Konzepte wie z.B. die Arbeitsmarktregionen umzusetzen und darüber hinaus auf der Basis individuell festgelegter Aggregationen vergleichende Analysen für Regionen mit unterschiedlichen wirtschaftlichen, soziodemografischen und sozioökonomischen Strukturen durchzuführen.

Bei der Anwendung des Hochrechnungsverfahrens auf regional tiefer Ebene muss berücksichtigt werden, dass nicht alle MZKR separate Hochrechnungsfaktoren für Deutsche und Ausländer aufweisen. Das hat für kleine Kreise mit niedrigem Ausländeranteil Abweichungen im Bezug auf die Ergebnisse der Bevölkerungsfortschreibung sowie auf höher aggregierter Ebene auch Abweichungen zu den bisher veröffentlichten Ergebnissen des Mikrozensus zur Folge. Dieses resultiert aus der Einführung der 10-Fälle-Grenze im Hochrechnungsverfahren für MZKR. Für die ostdeutschen Bundesländer werden keine gesonderten Hochrechnungsfaktoren für ausländische und deutsche Männer bzw. Frauen ausgewiesen, da hier die 10-Fälle-Grenze in den meisten MZKR unterschritten wird. Es wird aber auch für MZKR mit höherem Ausländeranteil auf separate Hochrechnungsfaktoren verzichtet, da für das ostdeutsche Gebiet ebenfalls keine getrennten Hochrechnungsfaktoren auf Anpassungsschichtebene für Deutsche und Ausländer ausgewiesen werden und so extreme Abweichungen zwischen den bisherigen Ergebnissen des Mikrozensus und den Ergebnissen auf MZKR-Ebene vermieden werden.

Die Betrachtungen zum Stichprobenfehler haben gezeigt, dass der relative Standardfehler bei einer Zellbesetzung von unter 50 einen Wert von über 15% annimmt. Dies gilt sowohl für Analysen im Grundfile als auch im Regionalfile. Begrenzender Faktor für die Durchführung von Punktschätzungen ist somit die Fallzahl einer spezifischen Gruppe und nicht die Größe der Regionaleneinheit an sich. Für die Wissenschaft ist der Mikrozensus aber nicht nur in Bezug auf Populationsschätzungen interessant. In mindestens ebenso großem Ausmaß ist das Forschungsinteresse auf Zusammenhangsanalysen bzw. die statistische Modellierung gerichtet. In diesem Kontext ist der Stichprobenfehler dann von untergeordnetem Interesse, da der Standardfehler ( $\beta$ ) im Modell nicht vom Auswahlverfahren, sondern von der Modellierung abhängt.

Mit den vorgenommenen Anonymisierungsmaßnahmen wird dem Datenschutz Rechnung getragen, bei zugleich größtmöglichem Erhalt des Analysepotenzials der Daten. Insbesondere durch die Bereit-

stellung der – entsprechend der Anonymisierungsvorgaben - unterschiedlich stark aggregierten Merkmale Beruf, Wirtschaftszweig, Alter und Nationalität kann die Regionalforschung – trotz gewisser Restriktionen – das breite Analysepotenzial des Mikrozensus für die Bearbeitung der spezifischen Forschungsfragen flexibel nutzen.

Auf der Grundlage des vorliegenden Papiers wurde ein erstes SUF für den Mikrozensus-Regionalfiler für die Off-Site-Nutzung entwickelt<sup>13</sup>. Hier wurde ebenfalls überprüft, inwieweit durch die Ergebnisse der 85% Sub-Stichprobe mit den Ergebnissen auf Basis des 100% Materials vergleichbar sind.

Neben dem SUF für die Off-Site-Nutzung kann das Mikrozensus-Regionalfiler auch für die Nutzung am Gastwissenschaftlerarbeitsplatz und über die kontrollierte Datenfernverarbeitung analysiert werden.

---

<sup>13</sup> Eine Beschreibung der Umsetzung des Anonymisierungskonzepts für das Jahr 2000 befindet sich im Dokument: Umsetzung der Anonymisierung.

## 6 Literatur:

Bihler, Wolf (2006): Stichprobendesign und Hochrechnung des Mikrozensus. ZUMA –Workshop. 9. März 2006, Mannheim.

Brauns, Hildegard; Steinmann, Susanne; Haun, Dietmar (2000): Die Konstruktion des Klassenschemas nach Erikson, Goldthorpe und Portocarero (EGP) am Beispiel nationaler Datenquellen aus Deutschland, Grossbritannien und Frankreich. ZUMA-Nachrichten (46), S. 7-63.

Hausen, Cornelia; Jungblut, Jean-Marie, Müller, Walter; Pollak, Reinhard und Heike Wirth (2006): Validation of ESeC: The Effect of Coding Procedures and Occupational Aggregation Level. Mannheim: 2006. [http://www.mzes.uni-mannheim.de/publications/papers/Validation\\_german\\_paper\\_lisboa.doc](http://www.mzes.uni-mannheim.de/publications/papers/Validation_german_paper_lisboa.doc)

Heidenreich, Hans-Joachim (1994): Hochrechnung im Mikrozensus ab 1990; in: Gabler, Siegfried u. a.: Gewichtung in der Umfragepraxis, Opladen, S. 112-123.

Meyer, Kurt (1994): Zum Auswahlplan des Mikrozensus ab 1990; in: Gabler, Siegfried u. a.: Gewichtung in der Umfragepraxis, Opladen, S.106-111.

Müller, Alexander (1992): Regionalauswertungen aus dem Mikrozensus, Neuer Auswahlplan führt ab 1990 zu verbesserten regionalen Auswertungsmöglichkeiten der 1%-Pepräsentativerhebung; in: Statistische Rundschau Nordrhein-Westfalen Heft 4, S. 179-189.

Müller, Walter u. a. (1991): Die faktische Anonymität von Mikrodaten; Band 19 der Schriftreihe Forum der Bundesstatistik herausgegeben vom Statistischen Bundesamt; Stuttgart.

Müller, Walter; Wirth, Heike; Bauer, Gerrit; Pollak, Reinhard; Weiss, Felix (2007): Entwicklung einer europäischen sozioökonomischen Klassifikation. In: Wirtschaft und Statistik, 5 (2007), S. 527-530

Rendtel, Ulrich; Schimpl-Neimanns, Bernhard (2001): Die Berechnung der Varianz von Populationschätzern im Scientific Use File des Mikrozensus ab 1996, ZUMA-Nachrichten (48), S. 85-116.

Schimpl-Neimanns, Bernhard ; Müller, Jörg Müller (2001): Zur Berechnung des Stichprobenfehlers im Mikrozensus. Mikrodaten-Tools 2001/06. Mannheim: ZUMA.

Statistisches Bundesamt (2000): Bevölkerung und Erwerbstätigkeit, Haushalte und Familien (Ergebnisse des Mikrozensus), Fachserie1, Reihe 3.

Statistisches Bundesamt (1999): Bevölkerung und Erwerbstätigkeit, Haushalte und Familien (Ergebnisse des Mikrozensus), Fachserie1, Reihe 3.

Statistisches Bundesamt (1997): Aufbereitung für den Mikrozensus 1997; III Kompensation der Ausfälle/Anpassung und Hochrechnung.



Statistisches Bundesamt (1996): Bevölkerung und Erwerbstätigkeit, Haushalte und Familien (Ergebnisse des Mikrozensus), Fachserie1, Reihe 3

Werner, Joachim (1994): Regionalisierung des Mikrozensus, Regionaler Bildausschnitt bei geringer Tiefenschärfe; in: Baden-Württemberg in Wort und Zahlen, Heft 6, S: 278-285.

Wirth, Heike; Schunck, Reinhard (2007): Zwischen Skylla und Charybdis: Die faktische Anonymisierung des Mikrozensus-Regionalfiles. ZUMA, Mannheim, April 2007 (unveröffentlichter Arbeitsbericht).

Wirth, Heike; Zühlke, Sylvia; Christians, Helga (2004): Der Mikrozensus als Datenbasis für die Regionalforschung; in: Grözing, Gerd; Matiaske, Wenzel (Hg.): Deutschland regional, Sozialwissenschaftliche Daten im Forschungsverbund; München und Mering, Rainer Hampp Verlag.

Wirth, Heike; Gresch, Cornelia; Müller, Walter; Pollak, Reinhard; Weiss, Felix (2009): Measuring social class: The case of Germany. In: David Rose, Eric Harrison (eds): Social Class in Europe. An introduction to the European Socio-economic Classification. Routledge. P.114-137.

**Anhang: 1**

Tabelle 2a: Ausgewählte Anpassungsschichten: Bevölkerung am Ort der Hauptwohnung, Bayern (absolute Häufigkeiten)

AS-Nr.	Mikrozensus ungewichtet				Mikrozensus gewichtet Kreisfaktor				Mikrozensus gewichtet Anpassungsschichtfaktor			
	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen
	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.	abs.
0906	2.189	2.315	102	87	237.324	249.420	16.778	12.045	238.000	247.800	16.400	14.400
0907	2.789	2.863	69	60	310.200	322.789	10.071	11.541	304.856	321.046	15.800	14.200
0908	2.370	2.368	74	69	246.502	255.567	7.659	8.971	244.424	255.776	10.000	9.100
0913	2.197	2.362	89	80	242.566	253.931	18.065	15.584	247.800	257.001	14.700	13.500
0918	1.972	2.046	71	64	214.355	223.043	9.856	8.410	212.772	221.928	11.300	9.700

Quelle: Mikrozensus 2000

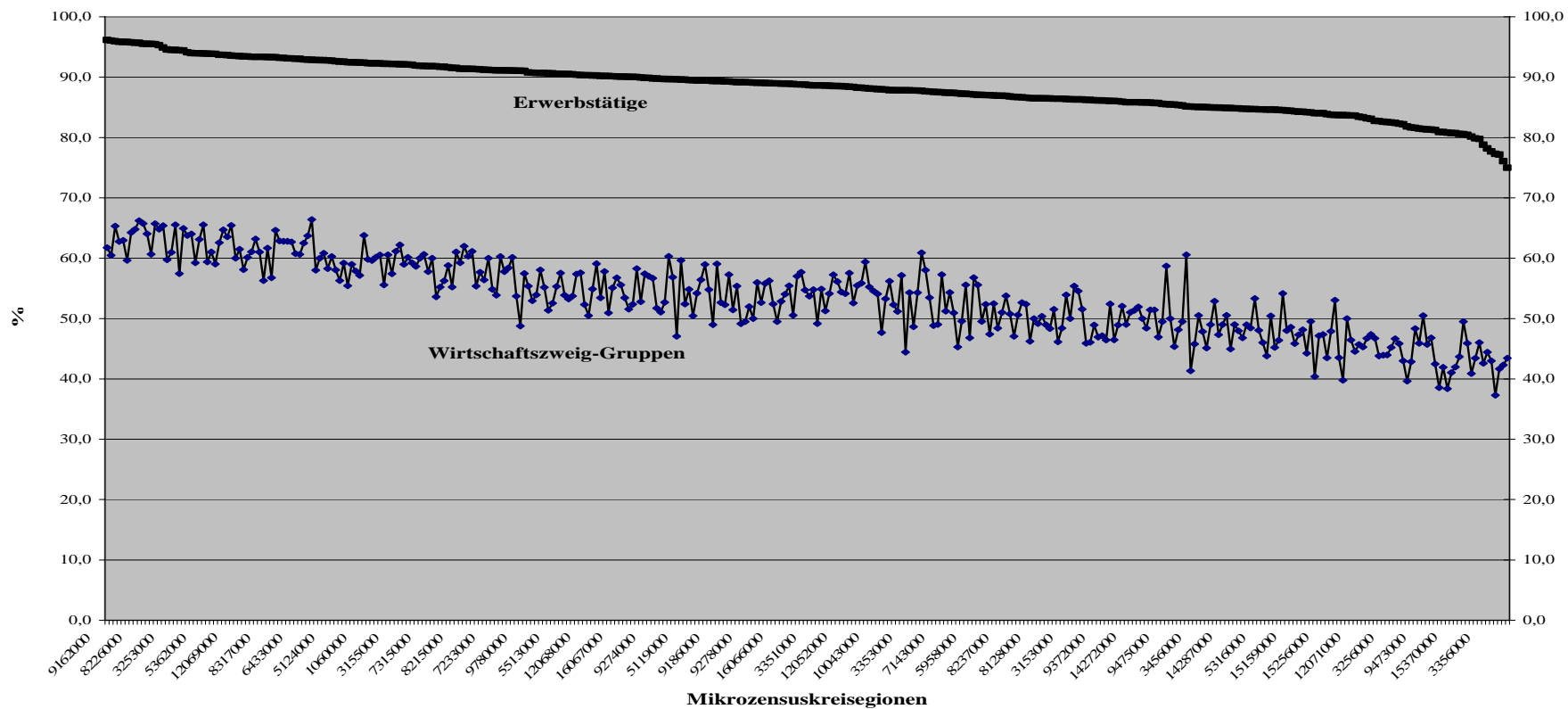
Tabelle 3a: Ausgewählte Anpassungsschichten/Kreise: Bevölkerung am Ort der Hauptwohnung; Bayern (in Prozent)

AS-Nr.	Mikrozensus ungewichtet				Mikrozensus gewichtet Kreisfaktor				Mikrozensus gewichtet Anpassungsschichtfaktor			
	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen	deutsche Männer	deutsche Frauen	ausländ. Männer	ausländ. Frauen
	%	%	%	%	%	%	%	%	%	%	%	%
0906	46,6	49,3	2,2	1,9	46,0	48,4	3,3	2,3	46,1	48,0	3,2	2,8
0907	48,2	49,5	1,2	1,0	47,4	49,3	1,5	1,8	46,5	48,9	2,4	2,2
0908	48,6	48,5	1,5	1,4	47,5	49,3	1,5	1,7	47,1	49,3	1,9	1,8
0913	46,5	50,0	1,9	1,7	45,8	47,9	3,4	2,9	46,5	48,2	2,8	2,5
0918	47,5	49,3	1,7	1,5	47,0	48,9	2,2	1,8	46,7	48,7	2,5	2,1

Quelle: Mikrozensus 2000

## Anhang 2

**Abb. 1: Anteil der nach Anonymisierung auf Bundes- und MZKR-Ebene verbleibenden Wirtschaftszweig-Gruppen (bezogen auf jeweils alle in einer MZKR besetzten Wirtschaftszweig-Gruppen) und der entsprechende Anteil an der erwerbstätigen Bevölkerung nach Mikrozensuskreisregionen**



**Abb. 2: Anteil der nach Anonymisierung auf Bundes- und MZKR-Ebene verbleibenden Wirtschafts-Abteilungen (bezogen auf jeweils alle in einer MZKR besetzten Wirtschafts-Abteilungen) und der entsprechende Anteil an der erwerbstätigen Bevölkerung nach Mikrozensuskreisregionen**

