

**Regionale Mortalitätsunterschiede  
in  
Baden-Württemberg**

Hans-Martin von Gaudecker

**46-2004**

Mai 2004

# REGIONALE MORTALITÄTSUNTERSCHIEDE IN BADEN-WÜRTTEMBERG

Hans-Martin von Gaudecker

Mannheimer Forschungsinstitut Ökonomie und demographischer Wandel (MEA)

Mai 2004

Diese Arbeit untersucht die zum Teil erstaunlich hohen Mortalitätsunterschiede in Baden-Württemberg. Zunächst wird die Möglichkeit ausgeschlossen, dass diese allein auf Zufallsschwankungen oder Fehlern im Meldewesen beruhen. Es schließen sich einige methodische Überlegungen an, bevor eine Ursachenanalyse durchgeführt wird. Als Hauptklärungsgrund für die Sterblichkeitsunterschiede konnte der sozioökonomische Status herausgearbeitet werden: In Kreisen mit hohem Einkommen leben die Menschen im Durchschnitt länger als in Kreisen mit geringem Einkommen. Der Einfluss des sozioökonomischen Status auf die Mortalität scheint durch höhere Bildung verstärkt oder sogar verursacht zu werden. Luftbelastung und Gesundheitsversorgung konnten nicht als diskriminierende Faktoren der Sterblichkeit in Baden-Württemberg festgestellt werden.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Mortalitätsdaten und deskriptive Analyse</b>	<b>3</b>
2.1. Beschreibung der Sterblichkeits- und Wanderungsdaten . . . . .	3
2.2. Berechnung der Kennziffern . . . . .	7
2.3. Darstellung der Mortalitätsunterschiede . . . . .	9
<b>3. Validierung der Ergebnisse</b>	<b>14</b>
3.1. Simulation von Zufallsschwankungen . . . . .	14
3.2. Simulation von Messfehlern . . . . .	18
<b>4. Theoretische Überlegungen</b>	<b>23</b>
4.1. Erklärungen für Mortalitätsunterschiede . . . . .	23
4.2. Ökonometrische Analyse aggregierter Daten . . . . .	27
4.3. Ein einfaches Regressionsmodell . . . . .	34
<b>5. Regressionsanalyse mit Paneldaten</b>	<b>39</b>
5.1. Datensatzbeschreibung und Hypothesenformulierung . . . . .	39
5.2. Ergebnisse . . . . .	48
<b>6. Zusammenfassung und Ausblick</b>	<b>57</b>
<b>A. Tabellen</b>	<b>59</b>
<b>Abkürzungsverzeichnis</b>	<b>89</b>
<b>Literaturverzeichnis</b>	<b>93</b>

## *Inhaltsverzeichnis*

# 1. Einleitung

Die menschliche Lebenserwartung bietet ein faszinierendes Forschungsfeld, weil sie von unzähligen Faktoren der verschiedensten Disziplinen beeinflusst wird. Individuen treffen jeden Tag sterblichkeitsrelevante Entscheidungen, die von ihrem sozialen und ökonomischen Kontext bestimmt sind. Die Mortalität hängt ebenfalls zusammen mit ihren genetischen Voraussetzungen, dem Stand medizinischen Wissens, dem Zugang zum Gesundheitssystem und den Umweltbedingungen, um nur einige Beispiele zu nennen. In dieser kurzen und weit von der Vollständigkeit entfernten Liste klingen bereits viele Fragestellungen der Wirtschafts- und Sozialwissenschaften, der Biologie, Medizin und Epidemiologie an. Die Lebenserwartung ist von höchster Relevanz sowohl für die individuelle Lebensplanung, als auch für die Organisation vieler gesellschaftlicher Institutionen, wie der Ausgestaltung des Rentensystems, der Krankenversicherung oder des Bedarfs an Pflege- und Altenheimen.

Besonders spannend ist der Alterungsprozess, der sich in den Industrienationen seit dem Ende des zweiten Weltkriegs scheinbar unaufhaltsam fortsetzt. Auch dieser lässt sich wiederum in unterschiedlichen Dimensionen betrachten. Die Längsschnittbetrachtung stellt die Frage, warum ein heute neugeborenes Kind eine höhere Lebenserwartung hat als seine Eltern, als diese auf die Welt kamen. In der Querschnittsbetrachtung wird nach Erklärungen dafür gesucht, dass zwei am selben Tag geborene Personen unterschiedliche Sterbewahrscheinlichkeiten aufweisen. Selbstredend ergänzen und befruchten sich beide Herangehensweisen gegenseitig.

Die Demografie nähert sich beiden Fragestellungen auf aggregierter Ebene und fragt nach den Sterblichkeitsunterschieden zwischen Bevölkerungen. In den Wirtschafts- und Sozialwissenschaften ergibt sich in der Frage nach der geeigneten Analyseebene ein Abwägen zwischen verschiedenen Übeln.

Auf individueller Ebene sind sozioökonomische Indikatoren in vielen seit den achtziger Jahren erhobenen Paneldatensätzen sehr detailliert enthalten, meist fehlt jedoch eine Kontrolle für den körperlichen Zustand der befragten Personen. Da sozioökonomische Entscheidungen eng mit diesem zusammen hängen, können sich insbesondere für alte Menschen große Verzerrungen ergeben. Allerdings ist gerade die Sterblichkeit dieser Bevölkerungsgruppe in der momentanen Phase des Alterungsprozesses von entscheidender Bedeutung.

## 1. Einleitung

Erst in letzter Zeit erhobene Paneldatensätze wie der HRS und AHEAD in den Vereinigten Staaten oder der europäische SHARE lassen eine ausreichende Kontrolle für den Gesundheitszustand zu. Hier liegt jedoch noch keine genügend große Zeitdimension vor, als dass fundierte Rückschlüsse auf die Lebenserwartung möglich wären.

Auf der Aggregatsebene hingegen kann aufgrund des Gesetzes der großen Zahl gehofft werden, dass die körperlichen Zustände über die Bevölkerungen hinweg weitgehend homogen sind. Es stellt sich die Frage, ob sich trotz der weniger scharfen Informationen zu sozioökonomischen Variablen beobachtbare Regelmäßigkeiten feststellen lassen, die Mortalitätsunterschiede erklären können. Falls ein signifikanter Einfluss von kontextuellen Faktoren wie dem Umwelteinfluss oder der Gesundheitsversorgung nachgewiesen werden kann, so hätten die Ergebnisse vor dem Hintergrund des Grundsatzes der Chancengleichheit auch direkt politikrelevante Konsequenzen.

In der vorliegenden Arbeit wurde der Kreis als Analyseeinheit für eine Querschnittsbetrachtung gewählt. Kapitel 2 stellt den verwendeten Datensatz vor und enthält eine knappe Beschreibung der regionalen Mortalitätsunterschiede in Baden-Württemberg. Die bloße Feststellung von Unterschieden in der Lebenserwartung hat vor dem Hintergrund endlich großer Populationen keine statistische Aussagekraft, wenn das Sterblichkeitsgeschehen stochastischer Natur ist. Mit der statistischen Validierung der differenziellen Mortalität beschäftigt sich das dritte Kapitel. Zunächst wird die Sterblichkeit als ein stochastischer Prozess modelliert und mittels Simulationen gezeigt, dass systematische Unterschiede der Lebenserwartungen vorliegen. Die Möglichkeit, dass Messfehler deren Ursache darstellen, wird ebenfalls über einen Simulationsansatz ausgeschlossen. Nachdem die Befunde auf diese Weise untermauert sind, wird im Rest der Arbeit versucht, Einflussfaktoren der Mortalität herauszuarbeiten.

Kapitel 4 enthält zunächst einen Überblick über die theoretische Literatur zu den Determinanten der Sterblichkeit. Ferner wird auf die Möglichkeiten und Grenzen der Analyse aggregierter Daten eingegangen. Abschließend findet sich ein Regressionsmodell, welches in Kapitel 5 mit Daten des Statistischen Landesamts Baden-Württemberg zur Anwendung kommt. Dieses bildet den eigentlichen Schwerpunkt der Arbeit und enthält verschiedene Modellspezifikationen zum Einfluss von sozioökonomischen Charakteristika, der Wanderungsbewegungen, des Gesundheitswesens, der Siedlungsstruktur sowie der Umwelt. Die Resultate zeigen deutlich den herausragenden Einfluss des sozioökonomischen Status auf die Lebenserwartung. Einkommen und höhere Bildung weisen eine signifikant negative Korrelation mit der Sterblichkeit auf. Zu den restlichen Einflussfaktoren finden sich gemischte Ergebnisse. Die Arbeit endet mit einer Zusammenfassung und einigen Schlussfolgerungen.

## 2. Mortalitätsdaten und deskriptive Analyse

Dieses Kapitel stellt die verwendeten Daten kritisch dar und liefert eine Beschreibung der nach Kreisen und Jahren differenzierten Sterblichkeit. Weiterhin wird auf die Methodik eingegangen, welche zur Berechnung der zugrundeliegenden Mortalitätskennziffern gewählt wurde.

### 2.1. Beschreibung der Sterblichkeits- und Wanderungsdaten

**Datenquelle** Sämtliche für diesen Abschnitt relevanten Daten wurden vom Statistischen Landesamt Baden-Württemberg zur Verfügung gestellt. Sie beinhalten für jeden der 44 Kreise des Bundeslandes den Bevölkerungsstand zum Jahresende von 1980 bis 2001, die innerhalb des betreffenden Jahres Gestorbenen, sowie die Wanderungsbewegungen mit Ziel bzw. Herkunft (bei Wanderungen innerhalb Baden-Württembergs den Kreis, ansonsten eine Untergliederung nach Deutschland / Ausland). Alle Daten sind aufgeschlüsselt nach Alter, Geschlecht und Staatsangehörigkeit (Deutsche / Ausländer). Die mindestens 100 Jahre alten Personen werden in einer Gruppe zusammengefasst. Auf einer höheren Aggregationsebene werden diese Daten regelmäßig veröffentlicht und sind im Wesentlichen verfügbar unter Statistisches Landesamt Baden-Württemberg (2003).

Die Daten sind insgesamt von hervorragender Qualität, perfektes Zahlenmaterial ist jedoch immer nur eine Wunschvorstellung und so gilt es, vor der Analyse einige wichtige Aspekte zu beachten. Alle Daten sind anonymisiert, d. h. Charakteristika können immer nur auf Kreisebene erfasst, aber nicht einzelnen Personen zugeordnet werden. Für die in diesem Abschnitt vorgenommene Beschreibung der Sterblichkeitsverhältnisse ist dabei relevant, dass räumliche Wanderungsbewegungen nicht auf der Individualebene erfasst werden.



## 2. Mortalitätsdaten und deskriptive Analyse

Falls jemand seinen Wohnort wechselt und kurz darauf stirbt, wird der Sterbefall im Zielkreis erfasst und dort in die Statistik einbezogen. Dies ist nicht sinnvoll für eine geographische Beschreibung der Mortalität, aber aufgrund der Datenstruktur unumgänglich. Sämtliche hier durchgeführten Analysen beinhalten somit die Annahme räumlich-stationärer Bevölkerungen, die selbstverständlich nicht erfüllt ist und als unvermeidbare Approximation angesehen werden sollte. Insbesondere im Fall von Umwelteinflüssen als Kovariate muss diese Vorgehensweise kritisiert werden, da man nur bei einer langfristigen Exposition Wirkungen auf die Mortalität erwarten würde. Im Zusammenhang damit soll auch bereits an dieser Stelle vor dem ökologischen Fehlschluss gewarnt werden: Zusammenhänge, die für bestimmte Gruppen (hier: Kreisbevölkerungen) gefunden werden, müssen nicht zwangsläufig für alle dazugehörigen Individuen gelten (Robinson 1950). Abschnitt 4.2 geht auf dieses Problem näher ein.

Der Bevölkerungsbestand bezieht sich auf alle Personen, die in dem betreffenden Kreis mit ständigem Wohnsitz gemeldet sind. Alle Daten basieren auf den offiziellen Ergebnissen der Volkszählungen 1970 und 1987. Zu den dort ermittelten Beständen werden jährlich die Zugänge durch Geburten und Zuwanderungen addiert, sowie die Abgänge durch Tod und Abwanderungen subtrahiert. Mit der Problematik dieser Fortschreibung befassen sich die nächsten beiden Absätze. Alle nicht gesondert gekennzeichneten Informationen zum Datenerhebungsprozess entstammen Gesprächen mit Herrn Ivar Cornelius vom Statistischen Landesamt Baden-Württemberg.

**Natürliche Bevölkerungsbewegung** Hierunter versteht man die Zu- und Abgänge durch Geburten und Tod. Diese werden dem Statistischen Landesamt von den Standesämtern der Gemeinden gemeldet. Die Qualität der Daten zu Neugeborenen und Sterbefällen innerhalb der Bundesrepublik Deutschland ist als sehr hoch einzustufen, da in jedem Fall vom Standesamt eine Urkunde erstellt wird. Lediglich im Zeitraum bis einschließlich 1994 könnten unsystematische Fehler durch manuelle Weiterverarbeitung in den Daten vorhanden sein, da diese erst seit 1995 elektronisch aufbereitet an das Statistische Landesamt Baden-Württemberg geliefert werden.

Sterbefälle im Ausland werden in aller Regel keinen Eingang in den Datensatz finden. Zwar existiert ein offizieller Meldeweg, von diesem wird allem Anschein nach jedoch selten Gebrauch gemacht. Falls diese Annahme stimmt, werden zu wenige Sterbefälle erfasst und der Bevölkerungsbestand in darauffolgenden Jahren zu hoch ausgewiesen. Wenn die Wahrscheinlichkeit im Ausland zu sterben für verschiedene Subpopulationen unterschiedlich ist (hohe Werte würde man z.B. für mobile Personen oder Ausländer erwarten), so kommt es

## 2.1. Beschreibung der Sterblichkeits- und Wanderungsdaten

bei den Mortalitätsindikatoren zu systematischen Verzerrungen. Diesem Problem trägt das Messfehlermodell in Abschnitt 3.2 im Rahmen des Möglichen Rechnung.

**Räumliche Bevölkerungsbewegung** Die größte Fehlerquelle im Rahmen der Fortschreibung ist bei der Erfassung von Zu- und Fortzügen zu erwarten. Bei allen Wanderungsbewegungen innerhalb Deutschlands basieren sämtliche Daten auf den Meldungen der Zielgemeinde (vgl. Statistisches Landesamt Baden-Württemberg (2000)). Folglich gründen sich die Fortzugszahlen auf schwer überprüfbare Angaben der betreffenden Personen, in welche sich leicht Fehler einschleichen können. Tabelle 2.1 illustriert ein einfaches Beispiel: eine Person aus dem Ortenaukreis beginnt ihr Studium in Heidelberg, ohne sich umzumelden. Bei Beginn des Berufslebens in Stuttgart meldet sie sich dort an und gibt als Herkunftsort Heidelberg an, wo sie jedoch nie gemeldet war. In den vorliegenden Daten wäre von diesem Moment an eine Person zu viel im Ortenaukreis und eine Person zu wenig in Heidelberg geführt.

Tabelle 2.1.: Beispiel für Fehler im Meldewesen

Periode	Tätigkeit	Ort	Anmeldung	Angegebener Herkunftsort	Fehler
1	Schulbesuch	Ortenaukreis	ja	-	+1
2	Studium	Heidelberg	nein	-	-1
3	Anstellung	Stuttgart	ja	Heidelberg	keiner

In Statistisches Landesamt Baden-Württemberg (2000) werden weitere anschauliche Möglichkeiten für Fortschreibungsfehler beispielhaft aufgeführt. Es lässt sich festhalten, dass Anmeldungen eher korrekt und vollständig erfasst werden als Abmeldungen. Damit führt dieses System tendenziell zu einer Überschätzung des Bevölkerungsbestandes. Eine weitere potenzielle Quelle dafür entspringt dem kommunalen Finanzausgleich, welcher auf den Bestandsdaten des Statistischen Landesamtes beruht. Für jede erfasste Person erhält eine Gemeinde bis zu EUR 1175 (Werte von 2000, vgl. Statistisches Landesamt Baden-Württemberg (2000)). Damit bestünde für diese ein finanzieller Anreiz, den Bevölkerungsbestand als zu hoch auszuweisen.

Tabelle A.1 im Anhang führt die Jahresendbestände der deutschen und ausländischen Bevölkerung in Baden-Württemberg auf. Die Kreise haben dabei eine Größe von 48.622 (Stadtkreis Baden-Baden, 1984) bis 599.415 Einwohner (Stadtkreis Stuttgart, 1992). Die

## 2. Mortalitätsdaten und deskriptive Analyse

geringste Ausländerzahl findet sich mit 3.306 im Stadtkreis Baden-Baden (1980), die höchste im Stadtkreis Stuttgart mit 145.907 (1996). Einigen Aufschluss über fehlerhafte Daten vermögen die Zahlen in Tabelle A.2 zu geben. Hier sind die mindestens 65 Jahre alten Personen gegliedert nach Nationalität und Geschlecht für ganz Baden-Württemberg aufgeführt. Interessant in Bezug auf die Fortschreibung sind insbesondere die Werte vor und nach der Volkszählung 1987, deren absolute Differenz und prozentuale Veränderung in den letzten beiden Zeilen aufgeführt sind. Die starken Rückgänge bei der ausländischen Bevölkerung um rund ein Viertel sind konsistent mit der Hypothese, dass viele nach Beendigung des Berufslebens in ihr Heimatland zurückkehren ohne sich abzumelden. Ein ähnliches, jedoch nicht so stark ausgeprägtes, Bild ergibt sich bei den Bevölkerungsbeständen über alle Altersgruppen hinweg.

**Manuelle Eingriffe** Die gesamten Rohdaten sind Jahresendbestände. Zur Berechnung der aktuariellen Mortalitätsraten sind jedoch die in einer Periode Gestorbenen auf die in dieser Periode von der Risikobevölkerung verlebten Personenjahre zu beziehen (vgl. Preston, Heuveline, und Guillot (2000), S. 19). Bei einjährigen Altersintervallen bietet sich das Jahresmittel der Bevölkerung als Approximation für die verlebten Personenjahre an. (vgl. Preston, Heuveline, und Guillot (2000), S. 15). Die Approximation gilt exakt, wenn die Veränderung der Bevölkerung durch Geburten, Wachstum und Tod durch eine lineare Funktion gegeben ist.

Die überwiegende Mehrheit der Daten weist keine sichtbaren Fehler auf. In wenigen Fällen trifft dies nicht zu und manuelle Korrekturen wurden erforderlich. In 3189 Fällen (0,85 %) treten negative Bevölkerungsbestände auf, womit keine Berechnung von Mortalitätsraten möglich ist. Die Kreisbevölkerungen wurden hier gleich Null gesetzt. Der Fehler tritt seit 1999 nicht mehr auf, da das Statistische Landesamt einen Negativausgleich eingeführt hat, welcher seinerseits zu Verzerrungen in hohen Altersklassen führen kann. Die Korrektur geht aus von der Hypothese, dass in der Fortschreibung lediglich das Alter falsch erfasst wurde. Bei Fortzug oder Tod einer Person in einer nicht besetzten Altersklasse, wird in den fünf Jahrgängen darunter nach Beständen gesucht und die Bewegung dort erfasst. Dieses Vorgehen ist insbesondere im Rahmen der nach oben hin offenen Altersklasse zu kritisieren und wird zur Zeit einer Revision unterzogen.

Ferner war in 501 Fällen (0,13 %) die Zahl der Gestorbenen größer als der mittlere Bestand. Zwar kann dies beispielsweise auftreten, wenn alle Mitglieder einer Altersklasse im Verlauf des betreffenden Jahres sterben; die hohe Zahl von negativen Beständen legt jedoch nahe, dass auch hier Messfehler die gewichtigere Rolle spielen dürften. Die Formel zur Konversion von Mortalitätsraten in Sterbewahrscheinlichkeiten (s. u. auf S. 8) beinhaltet

ferner die Restriktion, dass die Mortalitätsrate kleiner oder gleich zwei ist, was nicht in immer gegeben war. In all diesen Fällen wurde die Zahl der Lebenden gleich der Zahl der Gestorbenen gesetzt, da erstgenannte eher falsch gemessen wurden. Eine feinere Justierung des Eingriffs erschien nicht gerechtfertigt.

## 2.2. Berechnung der Kennziffern

Aus den Rohdaten wurden zunächst einfache Mortalitätsraten gegliedert nach Altersklasse (x), Zeitpunkt (Jahr=j), Kreis (k) und Geschlecht (sex=s) berechnet:

$$m_{x,j,k,s} = \frac{\text{Gestorbene}_{x,j,k,s}}{\text{Personenjahre}_{x,j,k,s}}$$

In 281 Fällen (0,15%) waren Altersklassen in einem Kreis nicht besetzt, so dass keine Schätzung der Mortalitätsrate möglich war. Als einfache Approximation wurde diejenige der darunterliegenden Altersklasse verwendet. Angesichts der geringen Zahl des Auftretens erschien eine kompliziertere Näherung nicht notwendig.

Um sinnvolle Vergleiche anhand weniger Kennzahlen zu ermöglichen, muss in irgend einer Form über die Altersdimension aggregiert werden. In der Literatur werden zu diesem Zweck verschiedene Verfahren vorgeschlagen, im Wesentlichen fiktive (fernere) Lebenserwartungen sowie standardisierte Sterbeziffern (vgl. Esenwein-Rothe (1982), Kapitel 2.3, oder Mueller (2000), Abschnitt 1.2.3). Erstgenannte sind fiktiver Natur, da sie auf einer Periodensterbetafel beruhen, während die Lebenserwartung ein Kohortenkonzept ist. Sie haben somit *keinerlei* Interpretation im eigentlichen Sinne des Wortes, sondern sind nur als ein pragmatischer Indikator für die Sterblichkeit einer Bevölkerung zu einem gegebenen Zeitpunkt anzusehen. Beispielsweise fließt in die Berechnung der fiktiven Lebenserwartung bei Geburt im Jahr 2003 die Sterblichkeit eines Neugeborenen und die eines heute Sechzigjährigen ein. Letztere sagt aber nicht viel darüber aus, wie die Sterblichkeitsverhältnisse in sechzig Jahren sein werden, welche für die Lebenserwartung des Neugeborenen relevant sind. Theoretisch sauberer ist die standardisierte Sterbeziffer, welche von vornherein Lebende und Gestorbene im Querschnitt betrachtet. Konkret werden die in einer Altersklasse Gestorbenen mit einer vorher definierten Standardbevölkerung gewichtet. Selbst wenn diese sinnvoll gewählt wird, kann es im Zeitablauf zu starken Verzerrungen kommen (vgl. Mueller (2000), S. 40). Aus dem Grund wurde die fiktive Lebenserwartung bei Geburt vorgezogen. Gegen diese wird gelegentlich angeführt, dass Zufallsschwankungen in den unteren Altersklassen die Ergebnisse stark beeinflussen können, da sie sich in den Sterbetafeln fortpflanzen. Diese Kritik ist in der vorliegenden Untersuchung aus zwei Gründen

## 2. Mortalitätsdaten und deskriptive Analyse

wenig relevant. Einerseits ist die Zeitreihe lang genug, um „Ausreißerjahre“ zu identifizieren. Zweitens wird dieses Problem in Abschnitt 3.1 aufgegriffen und gezeigt, dass solche Schwankungen nicht für die beobachteten Unterschiede in den Mortalitätskennziffern verantwortlich sein können. Die standardisierte Sterbeziffer wurde auf der Grundlage der Gesamtbevölkerung Baden-Württembergs im Jahr 1991 berechnet, jedoch aus Gründen der Übersichtlichkeit nicht aufgeführt, zumal beide Kennziffern qualitativ gleichwertige Ergebnisse lieferten. Schließlich wurde die fernere fiktive Lebenserwartung im Alter von 65 Jahren berechnet. Dies stellt einerseits die Vergleichbarkeit der Ergebnisse mit einer Studie für Hessen sicher (Wittwer-Backofen 1999), andererseits sind in dieser Altersklasse praktisch nur noch natürliche Todesursachen relevant (vgl. z. B. Gröner (2002)), auf welche diese Analyse im Wesentlichen abzielt. Bei der Interpretation der Ergebnisse ist jedoch beachtenswert, dass die Bevölkerung über 65 Jahren bereits mit Selektionsproblemen behaftet ist. So belegen Studien einen besonders großen Effekt des Einkommens- und Bildungsniveaus auf die Sterblichkeit in der Klasse der 25 bis 64-jährigen (Kitagawa und Hauser 1972).

Lebenserwartungen sind Funktionen von Sterbetafeln. Um diese zu berechnen, wurde auf das Verfahren von William Farr zurückgegriffen (vgl. Esenwein-Rothe (1982), S. 248), da es von der amtlichen Statistik verwendet wird und somit die Vergleichbarkeit sichergestellt ist. Dafür müssen die Mortalitätsraten zunächst in jährliche bedingte Sterbewahrscheinlichkeiten umgewandelt werden. Die grundlegende Annahme dabei ist, dass Sterbefälle über das Jahr gleichverteilt sind ( $a_x = \frac{1}{2}$  für  $x > 0$ ). Dies führt zu der folgenden Formel (Herleitung z.B. bei Preston, Heuveline, und Guillot (2000), S. 43):

$$1 - p_{x,j,k,s} = \frac{m_{x,j,k,s}}{1 + 0,5 \cdot m_{x,j,k,s}} \quad \text{für } x > 0$$

Nach Preston, Heuveline, und Guillot (2000) liefert diese Daumenregel bei der hier gegebenen Datenstruktur sehr präzise Ergebnisse, abgesehen von der untersten Altersklasse (vgl. ebenda, S. 46). Für Neugeborene wurde aufgrund des Einflusses der Säuglingssterblichkeit und der resultierenden stark rechtsschiefen Verteilung der Sterbefälle eine modifizierte Formel verwendet. Die geschlechtsspezifischen Formeln wurden von Coale, Demeny, und Vaughen (1983) auf der Basis sehr detaillierter Sterbedaten berechnet und werden von Preston, Heuveline, und Guillot (2000), S. 48, für Datensätze wie den hier verwendeten empfohlen:

$$1 - p_{0,j,k} = \frac{m_{0,j,k}}{1 + (0,45 + 2,684 \cdot m_{0,j,k}) \cdot m_{0,j,k}}, \quad \text{für Männer}$$
$$1 - p_{0,j,k} = \frac{m_{0,j,k}}{1 + (0,53 + 2,8 \cdot m_{0,j,k}) \cdot m_{0,j,k}}, \quad \text{für Frauen}$$

### 2.3. Darstellung der Mortalitätsunterschiede

Damit ist die erwartete Zahl der Lebensjahre in jeder Altersklasse gegeben als:

$$L_{x,j,k,s} = p_{x,j,k,s} + (1 - p_{x,j,k,s}) \cdot a_{x,j,k,s}$$

Die fiktive Lebenserwartung ergibt sich nun als deren gewichtete Summe, wobei die Wahrscheinlichkeit, die Altersklasse  $x$  zu erreichen, die Gewichte bildet:

$$e_0(j, k, s) = \sum_{x=0}^{100} \left( \prod_{i=0}^x p_{i,j,k,s} \right) \cdot L_{x,j,k,s}$$
$$e_{65}(j, k, s) = \sum_{x=65}^{100} \left( \prod_{i=65}^x p_{i,j,k,s} \right) \cdot L_{x,j,k,s}$$

Die höchste Altersklasse wurde wie eine einjährige behandelt. Das von Preston, Heuveline, und Guillot (2000) vorgeschlagene Verfahren führte insbesondere bei Ausländern aufgrund von stellenweise sehr geringen in dieser Altersklasse gemessenen Mortalitätsraten zu Fantasiewerten ( $e_0 = 160$  und mehr). Proberechnungen mit höheren Werten für die im Alter 100 verbleibende Lebenserwartung ergaben keine wesentlichen Unterschiede, so dass dieses Vorgehen aus pragmatischen Gründen gerechtfertigt erscheint.

## 2.3. Darstellung der Mortalitätsunterschiede

In diesem Abschnitt soll die Analyse der differenziellen Mortalität durch eine kurze Darstellung der regionalen Unterschiede motiviert werden. Detailliertere Ausführungen jüngerer Datums finden sich bei Cischinsky (2000) und Gröner (2002). Ein Vergleich der Tabellen A.3 bis A.6 sowie der Abbildungen 2.1 und 2.2 zeigt, dass beide Kennziffern ähnliche regionale Muster der Mortalität liefern. Aus diesem Grund wird hier im Wesentlichen die fiktive Lebenserwartung bei Geburt thematisiert und kurz auf systematische Unterschiede eingegangen. Zunächst sei jedoch auf die Tabelle 2.2 verwiesen, welche die Größenordnung der Effekte kurz verdeutlicht. Die fiktive Lebenserwartung bei Geburt liegt bei Frauen um fünf bis sechs Jahre höher als bei Männern, wobei die Männer jedoch höhere prozentuale Verbesserungsraten aufweisen. Hier lässt sich ein leichter Aufholprozess feststellen. Die Änderungsraten von Mittelwert, Minimum und Maximum liegen für Frauen sehr dicht beieinander, so dass sich hier keine eindeutigen Konvergenz- oder Divergenztendenzen abzeichnen. Bei Männern findet sich leichte Konvergenz, die Veränderungsrate des Minimums liegt über derjenigen des Maximums. Im Alter ergibt sich ein ähnliches Bild. Die fernere fiktive Lebenserwartung ist für Frauen drei bis vier Jahre höher als bei Männern, die prozentualen Veränderungsraten sind im letzten Fall höher. Hier lässt sich für beide Geschlechter

## 2. Mortalitätsdaten und deskriptive Analyse

eine gewisse Konvergenz feststellen. Die stärkere Verbesserung der Lebenserwartung für Männer könnte auf kriegsbedingte negative Selektionseffekte zurückzuführen sein, welche im Zeitablauf abnehmen (Dinkel 1985). Allerdings ist weder klar, ob die Differenzen statistisch signifikant sind, noch ob andere Effekte, zum Beispiel biologischer Natur, eine Rolle spielen.

Tabelle A.3 beinhaltet die fiktive Lebenserwartung bei Geburt für Männer in den Jahren 1981 und 2001, das arithmetische Mittel der jährlichen prozentualen Veränderungen, sowie den Mittelwert aller 21 Jahre. Aufgrund der relativ homogenen Veränderungsraten ist ein Vergleich auf Basis der letzten Spalte am aufschlussreichsten. Die im Mittel höchsten Werte finden sich im Bodenseekreis und in den Landkreisen Tübingen, Böblingen, Esslingen und Breisgau-Hochschwarzwald. Die Schlusslichter bilden der Stadtkreis Mannheim, der Neckar-Odenwald-Kreis, die Landkreise Freudenstadt und Rastatt, sowie die beiden Karlsruher Kreise. Bei den Frauen (Tabelle A.4) ergibt sich ein ähnliches Bild. Der Bodenseekreis und die Landkreise Tübingen und Breisgau-Hochschwarzwald bilden auch hier die Spitzengruppe, gefolgt von den Stadtkreisen Ulm und Stuttgart. Die geringste Lebenserwartungen sind wiederum im Neckar-Odenwald-Kreis, im Stadtkreis Mannheim und in den Karlsruher Kreisen zu finden, hinzu kommt der Landkreis Heilbronn.

Ein Vergleich über die Indikatoren hinweg ergibt, dass die Stadtkreise bei der ferneren Lebenserwartung im Alter 65 (Tabellen A.5 und A.6) durchweg besser abschneiden als im Fall derjenigen bei Geburt. So ist die Sterblichkeit von Frauen über 65 Jahren in Stuttgart am geringsten, Ulm und Freiburg liegen auch unter den ersten fünf. Das Muster gilt auch auf allen weiteren Positionen und ebenso für Männer.

Beachtung verdient auch die große Differenz der fiktiven Lebenserwartung bei Geburt zwischen Ausländern und Deutschen, die in Tabelle A.7 dargestellt ist. Eine Übersterblichkeit der einheimischen Bevölkerung lässt sich theoretisch über die Theorie heterogener Populationen begründen (Vaupel und Yashin 1985), exzellente Diskussionen dieses Themas finden sich bei Wittwer-Backofen (1999) und Cischinsky (2000). Die Größenordnung von bis zu elf Jahren erscheint jedoch unplausibel, was durch den Vergleich der durchschnittlichen fiktiven Lebenserwartungen vor und nach der Volkszählung untermauert wird (ganz unten in Tabelle A.7). Die Daten suggerieren, dass die Sterblichkeit von Frauen im ersten Zeitraum geringer ist als im letzteren, bei Männern ist keine Veränderung festzustellen. Auch dies muss nicht zwangsläufig auf Messfehlern beruhen: so schwächt sich zum einen der positive Selektionseffekt für „alteingesessene“ Ausländer ab, zum anderen könnte sich die Zusammensetzung der zuwandernden Bevölkerung geändert haben (die Bürgerkriegsflüchtlinge aus dem ehemaligen Jugoslawien in den neunziger Jahren zeichneten sich möglicherweise nicht durch eine besonders robuste Gesundheit aus). Nichtsdestotrotz erscheinen

### 2.3. Darstellung der Mortalitätsunterschiede

Tabelle 2.2.: Zusammenfassung der fiktiven Lebenserwartungen

Fiktive Lebenserwartung bei Geburt, Männer				
<b>Kreis</b>	<b>1981</b>	<b>2001</b>	<b>jährliche Änderung (%)</b>	<b>Mittelwert</b>
Mittelwert	71,33	77,03	0,37	74,19
Minimum	68,93	75,08	0,41	72,51
Maximum	73,75	78,64	0,31	75,63
Standardabweichung	0,94	0,77		0,57

Fiktive Lebenserwartung bei Geburt, Frauen				
<b>Kreis</b>	<b>1981</b>	<b>2001</b>	<b>jährliche Änderung (%)</b>	<b>Mittelwert</b>
Mittelwert	77,76	82,66	0,29	80,43
Minimum	76,11	80,77	0,28	79,17
Maximum	79,40	83,80	0,26	81,35
Standardabweichung	0,74	0,63		0,49

Fiktive fernere Lebenserwartung im Alter 65, Männer				
<b>Kreis</b>	<b>1981</b>	<b>2001</b>	<b>jährliche Änderung (%)</b>	<b>Mittelwert</b>
Mittelwert	14,06	17,11	0,94	15,39
Minimum	12,89	16,02	1,04	14,63
Maximum	15,04	17,92	0,84	16,07
Standardabweichung	0,45	0,49		0,33

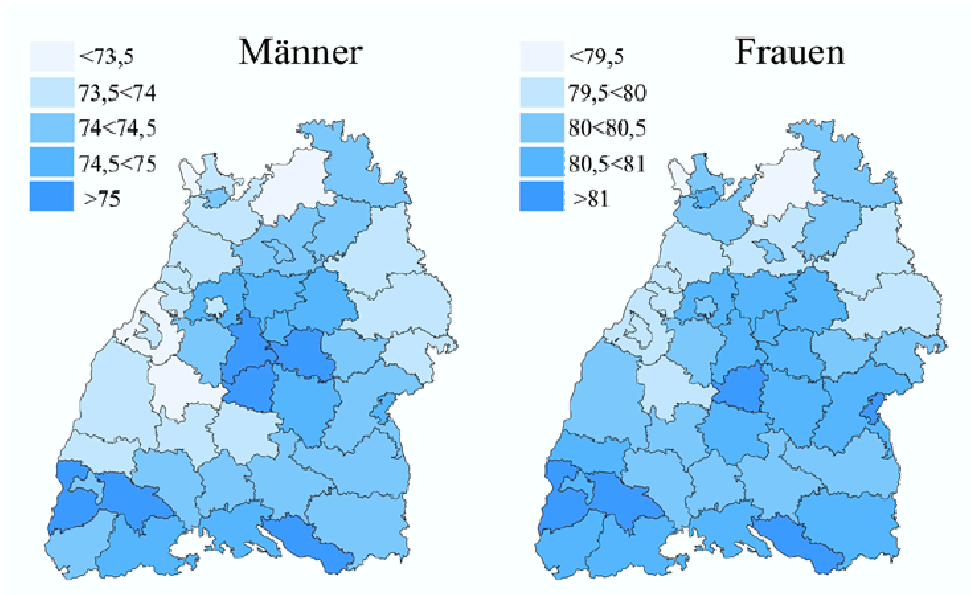
Fiktive fernere Lebenserwartung im Alter 65, Frauen				
<b>Kreis</b>	<b>1981</b>	<b>2001</b>	<b>jährliche Änderung (%)</b>	<b>Mittelwert</b>
Mittelwert	17,30	20,48	0,81	18,94
Minimum	16,14	19,29	0,85	17,81
Maximum	18,54	21,15	0,63	19,83
Standardabweichung	0,53	0,40		0,43

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.



## 2. Mortalitätsdaten und deskriptive Analyse

Abbildung 2.1.: Fiktive Lebenserwartung bei Geburt, Mittelwert 1981-2001.

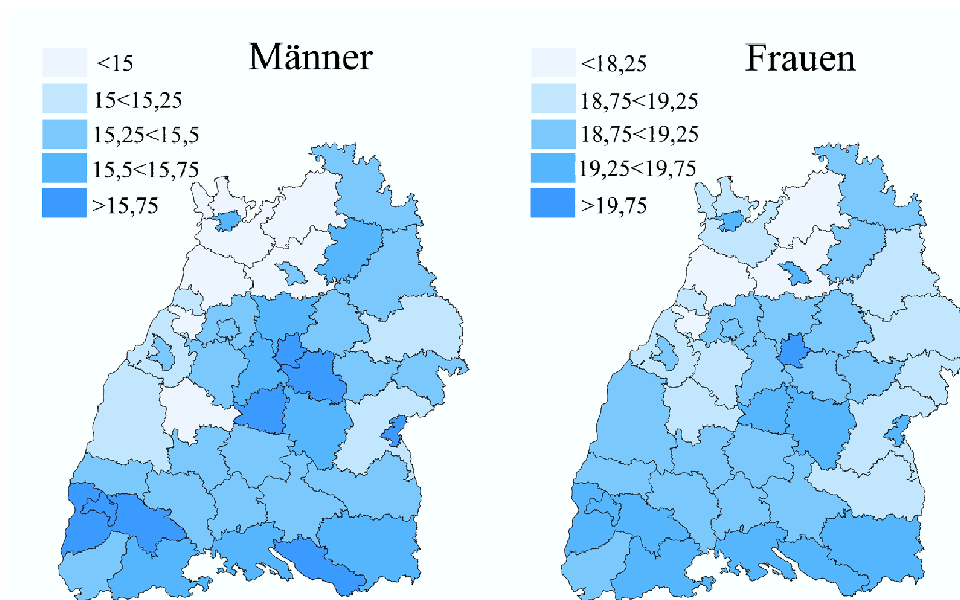


Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

die Differenzen sehr hoch und legen eine Überschätzung der Basisbevölkerung aufgrund von Problemen des Meldewesens nahe. Bei der fiktiven fernerer Lebenserwartung im Alter 65 treten diese Effekte noch deutlicher zu Tage, aufgrund der geringen Ausländerzahlen in dieser Altersklasse (s. Tabelle A.2) wurden diese jedoch nicht aufgeführt. Das im letzten Absatz angesprochene Muster der geringeren Alterssterblichkeit in den Stadtkreisen bleibt jedoch auch für die deutsche Bevölkerung bestehen und liegt nicht an höheren Ausländeranteilen.

### 2.3. Darstellung der Mortalitätsunterschiede

Abbildung 2.2.: Fiktive fernere Lebenserwartung im Alter 65, Mittelwert 1981-2001.



Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

## 3. Validierung der Ergebnisse

In diesem Kapitel soll untersucht werden, ob die Mortalitätsunterschiede notwendiger Weise auf regionalen Unterschieden beruhen. Dazu werden zwei alternative Möglichkeiten durchgespielt. Zum einen wird überprüft, ob Zufallsschwankungen aufgrund von kleinen Kreisbevölkerungen zu den beobachteten Unterschieden führen können. Zum anderen legt die oben aufgeführte Fortschreibungsproblematik die Möglichkeit systematischer Messfehler nahe, welche in Abschnitt 3.2 simuliert werden.

### 3.1. Simulation von Zufallsschwankungen

Dieser Abschnitt untersucht, ob die oben dargestellten Mortalitätsunterschiede daraus resultieren könnten, dass der Sterblichkeitsprozess stochastischer Natur ist und kleine Grundgesamtheiten in den Kreisen zu unterschiedlichen Realisationen führen. Zu diesem Zweck werden Zufallsziehungen aus einer Binomialverteilung simuliert, wobei die Sterbewahrscheinlichkeit im ganzen Land Baden-Württemberg als gleich angenommen wird. Anschließend werden die wahre und die simulierte Dispersion der fiktiven Lebenserwartungen verglichen. Wenn die wahre Streuung signifikant größer ist als die simulierte, so müssen andere Ursachen für die Mortalitätsunterschiede vorliegen. Eine zweite Möglichkeit zur statistischen Überprüfung wäre die direkte Berechnung von Konfidenzintervallen der fiktiven Lebenserwartungen (für eine Herleitung vgl. Smith (1992), Abschnitt 4.8, für eine Anwendung in Baden-Württemberg vgl. Scholz (2002)). Falls sich diese nicht in allen Fällen überlappten, so lägen statistisch signifikante Differenzen in der Sterblichkeit vor. Pro Jahr müsste man mit diesem Verfahren 44! Vergleiche auswerten und überprüfen, wie viele dieser Konfidenzintervalle sich überlappten. Mit dem Simulationsverfahren erhält man direkt eine Kennzahl, anhand derer man die Nullhypothese überprüfen kann. Letztgenannte lautet folgendermaßen:

$H_0$  : Die jahres- und geschlechtsspezifischen Kennziffern der Mortalität ( $e_0$ ,  $e_{65}$ ) sind in ganz Baden-Württemberg gleich.

### 3.1. Simulation von Zufallsschwankungen

Ausgehend davon wurden zunächst für jedes Jahr (1981-2001) und beide Geschlechter landesweite, altersspezifische Mortalitätsraten berechnet. Diese wurden mit oben beschriebenen Formelapparat in Sterbewahrscheinlichkeiten umgewandelt. Die Zahl der Gestorbenen einer Altersklasse jedes Kreises lässt sich nun als das Resultat eines stochastischen Prozesses begreifen, der einer Binomialverteilung folgt. Dessen Dichtefunktion ist dann gegeben als:

$$f = \frac{N!}{(N-X)! X!} p^{(N-X)} (1-p)^X$$

mit

$$f \sim U(0,1) \text{ simuliert mit einem Zufallszahlengenerator}$$

$$N = N(j, k, s, a) \text{ gemessene Zahl der Lebenden}$$

$$p = p(j, s, a) \text{ landesweit einheitliche Überlebenswahrscheinlichkeit}$$

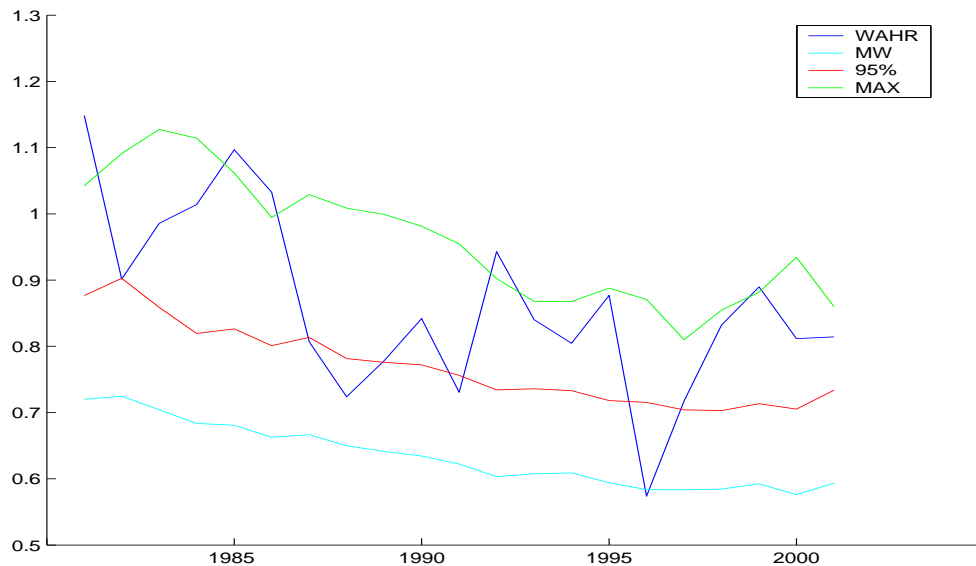
$$X = X(j, k, s, a) \text{ simulierte Zahl der Gestorbenen}$$

Die resultierenden Sterbefälle wurden zusammen mit den gemessenen Bevölkerungsbeständen dazu verwendet, die beiden fiktiven Lebenserwartungen zu berechnen. Insgesamt wurden 1000 Sterbetafeln für jedes Jahr und jeden Kreis berechnet. Die scheinbar kleine Zahl der Wiederholungen wird dadurch relativiert, dass die Indikatoren gewichtete Summen aus 101 bzw. 36 Ziehungen sind ( $e_0$  bzw.  $e_{65}$ ). Um die Kreisebene zu eliminieren, wird als Dispersionsmaß die Standardabweichung über die Kreise berechnet. Deren Verteilung in der Simulation wird nun mit den tatsächlichen Standardabweichungen verglichen. Wenn die real gemessenen Werte signifikant höher sind als die simulierten, so kann auf Unterschiede in der Sterblichkeit geschlossen werden, die nicht auf Zufallsschwankungen beruhen.

Die Ergebnisse finden sich in den Tabellen A.8 bis A.11 und den dazugehörigen Abbildungen 3.1 bis 3.4. In der ersten Spalte der Tabellen ist die empirische Standardabweichung der Kennziffern aufgeführt, in den weiteren der Mittelwert, das einseitige 95%-Konfidenzintervall und das Maximum der simulierten Werte. In der letzten Zeile werden diese Werte noch einmal über die Jahre gemittelt, um einen schnellen Überblick zuzulassen. Die Grafiken stellen diese Zeitreihen anschaulich dar. Für beide Geschlechter liegen die realen Standardabweichungen der fiktiven Lebenserwartung bei Geburt nur vereinzelt innerhalb des 95%-Konfidenzintervalls. Im Falle von  $e_{65}$  ist selbst dies nie gegeben, bei den Frauen liegt der empirische Wert sogar immer oberhalb des maximalen Simulationsergebnisses. Somit kann die Nullhypothese verworfen werden: die Mortalitätsunterschiede zwischen den einzelnen Kreisen sind nicht mit Zufallsschwankungen zu erklären.

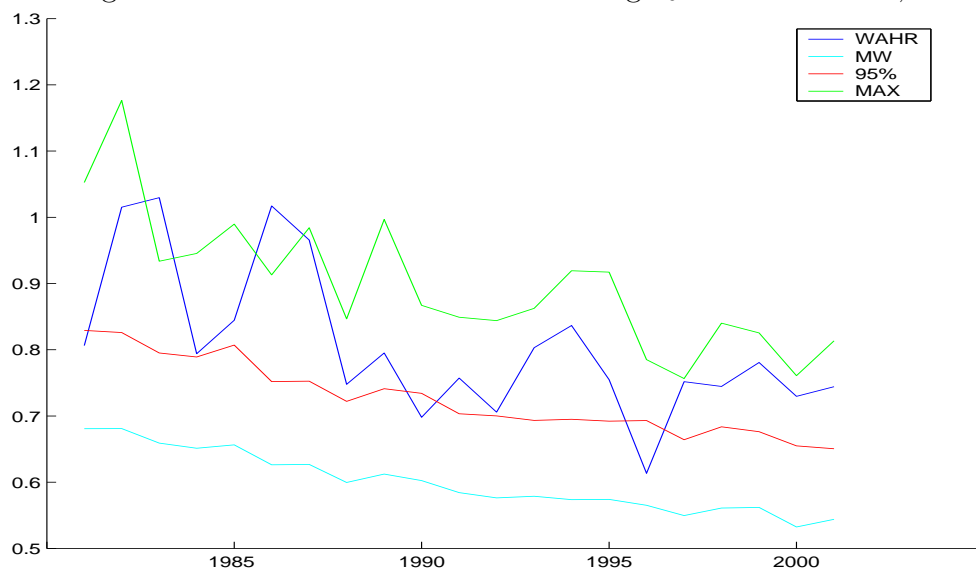
### 3. Validierung der Ergebnisse

Abbildung 3.1.: Simulation: Standardabweichung  $e_0$  über die Kreise, Männer



Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

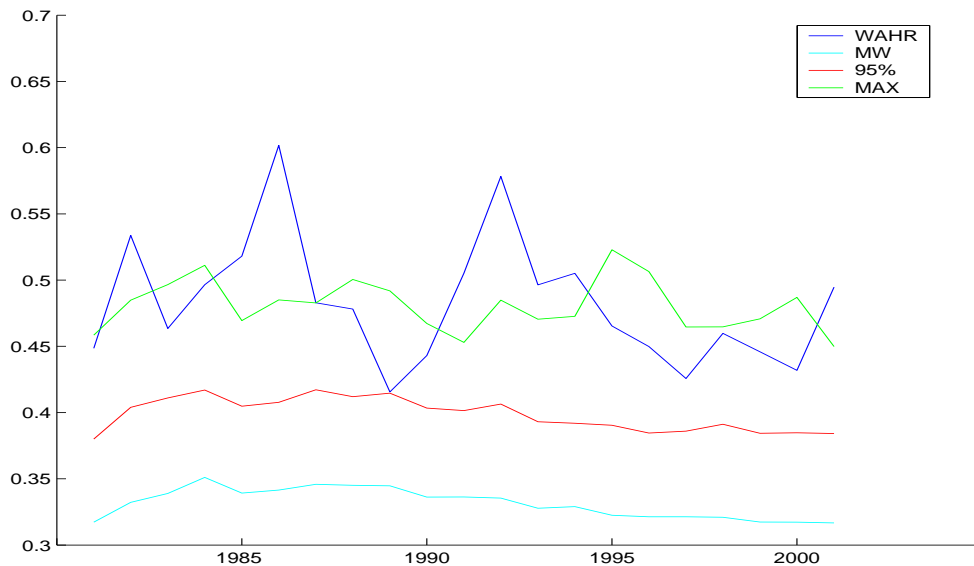
Abbildung 3.2.: Simulation: Standardabweichung  $e_0$  über die Kreise, Frauen



Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

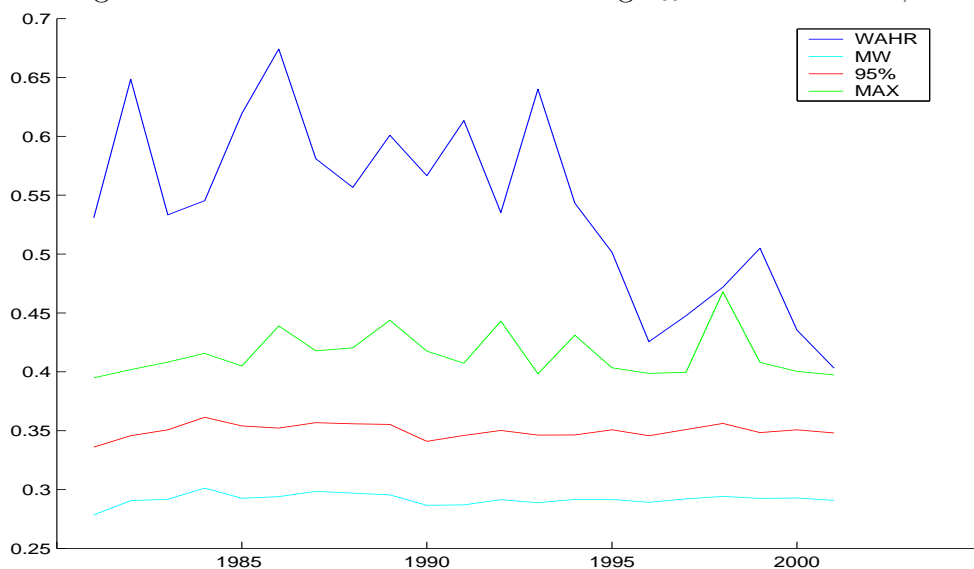
### 3.1. Simulation von Zufallsschwankungen

Abbildung 3.3.: Simulation: Standardabweichung  $e_{65}$  über die Kreise, Männer



Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Abbildung 3.4.: Simulation: Standardabweichung  $e_{65}$  über die Kreise, Frauen



Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

### 3. Validierung der Ergebnisse

**Zeitliche Varianz der kreispezifischen Sterblichkeit** Die Simulationsdaten wurden weiterhin verwendet, um zu überprüfen, ob Zufallsschwankungen die Variation der fiktiven Lebenserwartungen in einem Kreis über die Zeit erklären können. In diesem Fall lautet die Nullhypothese:

$H_0$  : Die Variation der kreis- und geschlechtsspezifischen Kennziffern der Mortalität ( $e_0$ ,  $e_{65}$ ) über die Zeit wird durch landesweit einheitliche Veränderungsraten und Zufallsschwankungen erklärt.

Der einzige Unterschied zur vorhergehenden Analyse besteht darin, dass die Standardabweichungen der fiktiven Lebenserwartungen nicht über die Kreise, sondern über die Zeit berechnet wurden. In den landesweit einheitlichen Mortalitätsraten sind jährliche Veränderungen enthalten. Wenn diese in einzelnen Kreisen größer ist, so sollten die Standardabweichungen höher sein und umgekehrt. Exemplarisch sind die Ergebnisse für die fiktive Lebenserwartung bei Geburt von Frauen in Tabelle A.12 aufgeführt. Die ersten beiden Spalten sind wiederum wahrer Wert und simuliertes Mittel, es folgen die obere und untere Grenze des beidseitigen 95%-Konfidenzintervalls. In drei Vierteln aller Fälle liegt die tatsächliche Standardabweichung innerhalb des Annahmebereichs. Die Nullhypothese kann somit nicht abgelehnt werden, was die Ergebnisse aus Abschnitt 2.3 untermauert. Trotz der Unzulänglichkeiten in der Methodik bleibt festzuhalten, dass der angenommene stochastische Prozess eine hinreichend gute Beschreibung des Sterblichkeitsgeschehens darzustellen scheint. Die Verbesserungsraten der Mortalität sind landesweit weitgehend einheitlich, die Sterblichkeitsunterschiede zwischen den einzelnen Kreisen sind im gesamten Zeitraum gegeben.

## 3.2. Simulation von Messfehlern

Wie bereits in Abschnitt 2.1 dargelegt wurde, sind die Bevölkerungsbestandsdaten eindeutig mit Fehlern behaftet. An dieser Stelle soll der Frage nachgegangen werden, ob sich die Mortalitätsunterschiede allein mit systematischen Messfehlern erklären lassen. Dazu wird ausgehend von Informationen zum Datenerfassungsprozess ein Simulationsmodell aufgebaut, welches in die entgegengesetzte Richtung des Messfehlers wirkt. Falls sich die Dispersion der fiktiven Lebenserwartungen reduziert, so kann auf einen signifikanten Einfluss von Messfehlern geschlossen werden.

Aus den oben aufgeführten Informationen zur Fortschreibungsmechanik (s. Abschnitt 2.1) wurden die folgenden stilisierten Fakten abgeleitet:

- Die Volkszählungsergebnisse sind fehlerfrei.
- Geburten und Todesfälle in Deutschland werden korrekt erfasst.
- Bei der Erfassung von Wanderungsbewegungen treten Fehler auf, die tendenziell zu einer Überschätzung des Bevölkerungsbestandes führen. Zuwanderungen werden häufiger korrekt erfasst als Abwanderungen.
- Die Fehler sind bei Ausländern gravierender als im Falle von deutschen Staatsangehörigen.

Diese vier Punkte wurden für das Simulationsmodell parametrisiert. Da keine geschlossene Zeitreihe für Sterblichkeits-, Bestands- und Migrationsdaten seit der Volkszählung 1970 vorlag, konnte die Schätzung nur für den Zeitraum von 1987 bis 2001 durchgeführt werden. Ausgehend von den Volkszählungsdaten wurde die Fortschreibung manuell durchgeführt, wobei die Wanderungsdaten mit einem normalverteilten Messfehler behaftet sind. Konkret sieht das Modell wie folgt aus:

$$\begin{aligned} \text{Bestand}_{t+1} &= \text{Bestand}_t + \text{Geburten}_t - \text{Sterbefälle}_t \\ &\quad + \text{Zuwanderungen}_t - \text{Abwanderungen}_t \\ &\quad - \text{Messfehler}_t \end{aligned}$$

$$\text{Messfehler}_t \sim N(E_{t,n}, V_{t,n})$$

$$E_{t,\text{Deutsche}} = \frac{1}{50} \cdot \text{Abwanderungen}_t + \frac{1}{500} \cdot \text{Zuwanderungen}_t$$

$$E_{t,\text{Ausländer}} = \frac{4}{50} \cdot \text{Abwanderungen}_t + \frac{4}{500} \cdot \text{Zuwanderungen}_t$$

$$V_{t,n} = 5 \cdot E_{t,n}$$

Mit anderen Worten, unter dieser Annahme wird jeder fünfzigste Fortzug von Deutschen wird erwartungsgemäß nicht erfasst, entsprechend ist jede fünfhundertste Zuwanderung in den Daten fehlerhaft. Die erwarteten Fehlergrößen vervierfachen sich für Ausländer, die Varianz ist linear-positiv mit dem erwarteten Fehler korreliert (d. h. je mehr Wanderungen anfallen, desto mehr Fehler treten auf; es bestehen sozusagen negative Skaleneffekte). Die Fehler kumulieren sich über die Zeit. Die Zahlenwerte wurden aufgrund von Plausibilitätsannahmen gewählt, selbstverständlich wurden auch andere Parametrisierungen getestet. Nicht direkt berücksichtigt sind Überschätzungen des Bevölkerungsbestandes, die durch

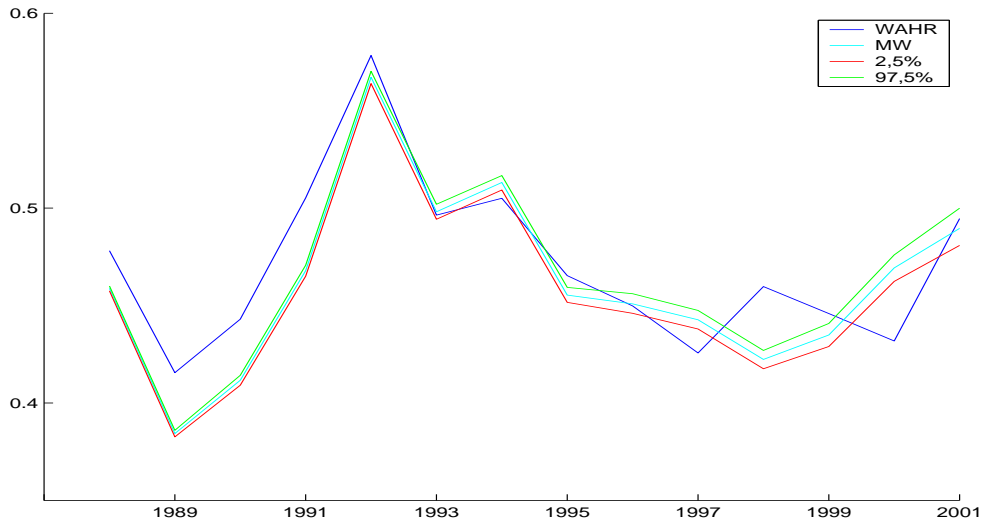


### 3. Validierung der Ergebnisse

nicht erfasste Todesfälle im Ausland zu Stande kommen. Der Fall, dass Ausländer nach Beendigung ihres Berufslebens in ihr Heimatland zurückkehren und dort sterben, wird jedoch approximativ von den erhöhten Erwartungswerten für Ausländer beim Bestandsfehler abgedeckt. Dies gilt exakt unter der plausibel erscheinenden Annahme, dass in Deutschland verweilende und in die Heimat zurückkehrende Ausländer dieselbe Lebenserwartung haben. Anders ausgedrückt, die Entscheidung zur Rückkehr ist unabhängig vom Gesundheitszustand. Alle anderen Todesfälle im Ausland entziehen sich einer systematischen Modellierung. Die Wahrscheinlichkeit eines Todes im Ausland würde sich zum Beispiel für sehr mobile Personen erhöhen - Daten zu Reisegewohnheiten sind jedoch nicht vorhanden.

Der Messfehler wurde verbal so dargestellt, wie er annahmegemäß auftritt. Da er jedoch in den Daten bereits vorhanden ist, wird er im Simulationsmodell von der Fortschreibungsgleichung subtrahiert und wirkt in die entgegengesetzte Richtung. Es wird also versucht, ausgehend von den fehlerhaften Daten der tatsächlichen Bevölkerung näher zu kommen. Da alle Aussagen aufgrund der willkürlichen Parametrisierung notwendiger Weise rein qualitativer Natur sind, können Basiseffekte vernachlässigt werden. Aus der simulierten Bevölkerung und den erfassten Todesfällen wurden neue Sterbetafeln berechnet und die Standardabweichung über die Kreise mit den gemessenen Werten verglichen. Ein Absinken wäre ein Indiz für einen positiven Einfluss von Messfehlern auf die Unterschiede in den fiktiven Lebenserwartungen.

Es wurde eine Reihe von verschiedenen Parametrisierungen durchgeführt, die hier beschriebene hat rein exemplarischen Charakter. Andere Spezifikationen führten entweder zu ähnlichen oder sehr unrealistischen Ergebnissen. Die Modelle wurden anhand von zwei Kriterien bewertet: zum einen die Veränderung der Standardabweichung der fiktiven Lebenserwartungen, zum anderen die Entwicklung der offensichtlichen Fehler als Anhaltspunkt für die Kalibrierung. Eine starke Erhöhung des Proportionalitätsfaktors für die Varianz führt zum Beispiel zu einer sehr hohen Zahl von negativen Beständen. Zwar würde man aufgrund des negativen Erwartungswertes des Fehlers im Simulationsmodell eine größere Steigerung im Vergleich zur Realität erwarten, zehnmal so hohe Werte wie die gemessenen erschienen dem Autor jedoch unplausibel. Das erste Kriterium war nie in einer Größenordnung erfüllt, die auf signifikante Einflüsse eines solchen Messfehlers schließen lässt. Bei der oben beschriebenen Spezifikation ließ sich ein leichter Rückgang nur für  $e_{65}$  bei Männern feststellen, in allen anderen Fällen erhöhte sich die Standardabweichung sogar oder blieb gleich. Die exemplarisch aufgeführten Ergebnisse finden sich in Tabelle A.13 und Abbildung 3.5. Auffällig ist, dass die simulierten Werte trotz des kumulierten Fehlers auch am Ende des Zeitraums nur ein sehr schmales Konfidenzintervall aufweisen und sich weitgehend im Einklang mit den gemessenen Werten bewegen.

Abbildung 3.5.: Simulation von Messfehlern: Standardabweichung  $e_{65}$  über die Kreise, Männer

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Aus den Ergebnissen kann geschlossen werden, dass ein Messfehler wie der hier angenommene keine Erklärung für die Sterblichkeitsunterschiede bieten kann, eher wirkt er in die entgegengesetzte Richtung. Obwohl vieles an der Methodik zu kritisieren bleibt, erscheint ein weiteres Vorgehen auf diesem Wege (zum Beispiel andere Verteilungsannahmen oder grundlegend verschiedene Spezifikationen von Fehlern für beide Staatsangehörigkeiten) aus mehreren Gründen wenig erfolgversprechend. Zum einen lässt sich in den Daten kein Muster erkennen, dass zum Zeitpunkt der Volkszählung eine deutlich geringere Dispersion der fiktiven Lebenserwartungen aufweist als in den Folgejahren. Dies sollte man aber erwarten, wenn sich die Messfehler im Zeitablauf fortpflanzen. Zum anderen liegt vor allem eine starke Überschätzung des Ausländerbestandes nahe. In den Kreisen mit hohen Ausländeranteilen haben die Indikatoren aber in aller Regel eine relativ geringe Ausprägung. Mannheim hat beispielsweise eine der niedrigsten fiktiven Lebenserwartungen und eine der höchsten Ausländerquoten. Würde der Bestand hier relativ stark überschätzt, so sanken die Indikatoren noch weiter und die Varianz über das Land hinweg stiege. Entsprechendes gilt für die meisten Stadtkreise.

Vorausgreifend sei bereits ein Hinweis auf Kapitel 5 erlaubt, wo sich aus einem anderen Ansatz heraus Anhaltspunkte finden, die gegen einen solchen Messfehler sprechen (vgl. S. 46). Abschließend lässt sich festhalten, dass Messfehler zweifellos auftreten, in den vorstellbaren Größenordnungen scheinen sie jedoch keinen systematischen Effekt auf die hier

### *3. Validierung der Ergebnisse*

betrachteten Sterblichkeitsmaße auszuüben. Die Suche nach „realen“ Gründen für die beobachteten Mortalitätsunterschiede ist damit endgültig gerechtfertigt.

## 4. Theoretische Überlegungen

Dieses Kapitel dient der theoretischen Fundierung des Hauptteils der Arbeit. Zunächst wird ein knapper Überblick über die theoretische und empirische Literatur zur Erklärung von Mortalitätsunterschieden gegeben. Anschließend wird auf die Analyse aggregierter Daten eingegangen und ein einfaches Regressionsmodell für die gegebene Datenstruktur entwickelt.

### 4.1. Erklärungen für Mortalitätsunterschiede

Im Wesentlichen finden sich in der Literatur zwei Herangehensweisen zur Erklärung der Sterblichkeit: die medizinisch-biologische und die sozialwissenschaftliche (Klein (1993), S. 712). In den Sozialwissenschaften wird dabei in der Regel die medizinisch-biologische Komponente als „black box“ verstanden, so dass beiden Herangehensweisen unterschiedliche Kausalitätsbegriffe zugrunde liegen (vgl. Theurl (1991), S. 186-188). Wittwer-Backofen (1999) beleuchtet in theoretischer Sicht beide Aspekte, kann die medizinisch-biologische Determinanten mit ihren Daten empirisch jedoch nicht berücksichtigen. Die Erhebung von sozioökonomischen Daten bei groß angelegten Studien zur medizinischen Wirksamkeit von Therapien ist noch keine gängige Praxis (Smith 1999), eine Ausnahme bildet die WHO-MONICA-Studie (Klein, Schneider, und Löwel 2001).

Die vorliegende Arbeit hält sich an den sozialwissenschaftlichen Ansatz. Für die ökonometrische Analyse der Daten ist entscheidend, ob die Regressoren im strukturellen Modell individuenspezifischer Natur sind oder kreisweite (kontextuelle) Faktoren darstellen (vgl. z. B. Greenland (2002), S. 390, Näheres in Abschnitt 4.2). Das strukturelle Modell auf individueller Ebene mit einem Mortalitätsindikator  $Y_{ik}$  als abhängiger Variable, individuellen Kovariaten  $X_{ik}$  und kreisweit einheitlichen Regressoren  $Z_k$  lässt sich für Kreis  $k$  und Individuum  $i$  als ein generalisiertes lineares Modell wie folgt schreiben:

$$(4.1) \quad E[Y_{ik}] = \mu_{ik} = f(\alpha + X_{ik}\beta + Z_k\gamma) .$$

#### 4. Theoretische Überlegungen

Anhand dieser Unterscheidung sollen nun die wesentlichen Faktoren dargestellt werden, welche aus theoretischer Sicht die Mortalität beeinflussen. Im Zuge dessen werden auch ausgewählte empirische Ergebnisse aufgeführt. Unter den individuellen Faktoren hat der sozioökonomische Status eine besondere Stellung in der Literatur erhalten, weswegen er zunächst gesondert behandelt wird. Die weiteren Kategorien umfassen sonstige individuelle Einflüsse und kreisspezifische Merkmale.

**Sozioökonomischer Status** Es gibt breite internationale Evidenz für einen negativen Zusammenhang zwischen sozialer Lage und Sterblichkeit (für einen Überblick vgl. Hummer, Rogers, und Eberstein (1998)). Insbesondere der Einfluss des Einkommens ist in der Literatur stark vertreten, ein umfassender Überblick findet sich bei Smith (1999), für knappere Darstellungen mit umfangreichen Literaturangaben vgl. Reil-Held (2002) oder Klein (2001). Direkte darauf aufbauende Effekte sind zum Beispiel Wohnverhältnisse oder der Zugang zu Einrichtungen des Gesundheitswesens (Klein 2001). Indirekte Wirkungshypothesen, die allerdings auch stark von mit dem Einkommen korrelierten Faktoren wie Bildung oder beruflicher Stellung zusammenhängen, gründen sich unter anderem auf Familienhintergrund, schichtspezifische Ernährungs- und Lebensgewohnheiten, Arbeitsplatzverhältnisse, Behandlung während der Schwangerschaft und in der frühesten Kindheit, sowie psycho-sozialen Stress durch Einkommensunterschiede in der Gesellschaft (Smith (1999), S. 157ff.). Während einige der genannten Hypothesen bestätigt werden konnten, besteht die Diskussion über die Größenordnung der Effekte unvermindert fort (Smith (1999), S. 145-146). Vor allem Ökonomen weisen immer wieder auf die entgegengesetzte Kausalität hin, d. h. vom Gesundheitszustand (der stark mit der Lebenserwartung korreliert ist) auf das Einkommen (für eine ausführliche Diskussion dieser Effekte vgl. Smith (1999)). Neuere Panelstudien mit US-Daten belegen eher schwach ausgeprägte Kausaleffekte in beide Richtungen (Hurd et al. 2003). Einen gegenteiligen Effekt machen Snyder und Evans (2002) aus, denen eine exogene Verringerung des Renteneinkommens in den USA als natürliches Experiment zur Verfügung steht. Die betroffene Kohorte reagierte mit einer moderaten Ausweitung des Arbeitsangebots im Rentenalter, was zu einer geringeren Sterblichkeit führte. Dies gilt natürlich nur für eine sehr kleine selektive Gruppe von Individuen, verdeutlicht allerdings die potentiellen Nichtlinearitäten und Interaktionen. Lleras-Muney (2002) weist für die USA einen signifikanten Kausaleffekt des Bildungsniveaus auf die Lebenserwartung nach, dort findet sich auch ein Überblick über die entsprechende Literatur. In Deutschland finden Klein, Schneider, und Löwel (2001) einen positiven Zusammenhang von Bildung und Lebenserwartung.

Für die Zwecke dieser Arbeit bleibt die Assoziation von sozialer Lage und Mortalität

#### 4.1. Erklärungen für Mortalitätsunterschiede

festzuhalten, weitergehende Überprüfungen der genannten Kausalitätsmechanismen sind mit dem gegebenen Datenmaterial nicht möglich. Der Einfluss sozioökonomischer Determinanten auf die Lebenserwartung hat in Deutschland in den letzten 15 Jahren zunehmende Beachtung gefunden (Klein 1993), eine positive Korrelation wurde auch für die Bundesrepublik eindeutig nachgewiesen (für Überblicke und Analysen jüngerer Datums auf Kohortenbasis vgl. Klein (2001) oder Reil-Held (2002)).

**Sonstige individuelle Faktoren** Ein tautologischer Einflussfaktor ist das Alter, welches positiv mit dem Mortalitätsrisiko korreliert ist. Ausnahmen bilden lediglich die Säuglingssterblichkeit und der sogenannte „Unfallknick“ bei jungen männlichen Erwachsenen (vgl. z. B. Gröner (2002)). Falls fiktive Lebenserwartungen oder standardisierte Sterbeziffern als abhängige Variablen verwendet werden, so ist eine weitere Beachtung des Altersaufbaus der Bevölkerung nicht notwendig. Dies kann jedoch zu ökonometrischen Problemen führen, wenn die erklärenden Variablen anderweitig standardisiert sind (Rosenbaum und Rubin (1984), s. a. Abschnitt 4.2). Weithin bekannt ist auch die männliche Übersterblichkeit, für einen knappen Überblick vgl. Anson (2003). Ihr wird Rechnung getragen, indem die Regressionen nach Geschlechtern getrennt durchgeführt werden.

Die medizinisch-biologische Literatur betont die genetische Ausstattung von Individuen, die hier jedoch nicht weiter verfolgt werden kann. Einerseits sind im Rahmen dieser Analyse keine Daten verfügbar, andererseits erscheint die Annahme homogener genetischer Dispositionen auf Kreisebene wenig restriktiv (vgl. Wittwer-Backofen (1999), S. 243-244). Damit zusammen hängt die ethnische Zugehörigkeit, die im Datensatz jedoch auch nur unzureichend erfasst ist. Die Differenzierung nach deutscher und ausländischer Staatsangehörigkeit ist nicht hinreichend, da viele Mitglieder ausländischer ethnischer Gruppen die deutsche Staatsangehörigkeit besitzen und die Zugewanderten eine sehr selektive Stichprobe darstellen (s. Abschnitt 2.3). Auch der Einfluss ethnischer Zugehörigkeit auf die Mortalität muss somit als kreisweise homogen angenommen werden. Ferner ist eine geringere Sterblichkeit von Verheirateten belegt, wofür es verschiedene Erklärungen gibt (Gärtner 1990). Die Ehegemeinschaft könnte eine „Kontrollfunktion“ ausüben, die zu einer gesünderen Lebensweise führt. Todesursachenvergleiche legen nahe, dass Verheiratete deutlich seltener an beeinflussbaren Todesursachen wie Lungenkrebs, Leberzirrhose oder Selbstmord sterben (Fuchs 1974). Die Hypothese, dass die Unterschiede daraus resultieren, dass nur Gesündere überhaupt heiraten, erscheint wenig stichhaltig, da die Sterberaten Verwitweter und Geschiedener in den meisten Altersgruppen höher sind als die von Ledigen (Dorbritz und Gärtner 1995). Leider liegen keine Daten zum Familienstand vor, wiederum muss eine Homogenitätsannahme auf Kreisebene gelten.

#### 4. Theoretische Überlegungen

Aufgrund der Datenverfügbarkeit sollten als reine Kontrollfaktoren auf individueller Ebene solche berücksichtigt werden, die als Indikatoren für den Gesundheitszustand taugen, obwohl sie aus theoretischer Sicht keinen eigenen Einfluss darauf ausüben. Dies gilt insbesondere für die Mobilität. Ein direkter Einfluss von Umzügen auf das individuelle Mortalitätsrisiko erscheint eher unwahrscheinlich, jedoch werden während des Berufslebens tendenziell gesündere Menschen umziehen. Unter dieser Annahme böten sich die Wanderungsbewegungen als ein Indikator der Sterblichkeitsverhältnisse an. Dieser Trend könnte sich im Rentenalter umkehren, wenn ältere Menschen aufgrund von Pflegebedürftigkeit in eine entsprechende Einrichtung oder zu ihren Kindern ziehen (Klein 1998). Der erstgenannte Selektionseffekt gilt, wie bereits in Abschnitt 2.3 angesprochen, für Ausländer in ähnlicher Form. Eine weitere wünschenswerte Kontrollvariable für kreisspezifisch unterschiedliche Zusammensetzungen der Bevölkerung wäre der Behindertenanteil (vgl. Cischinsky (2000), S. 66), da trotz homogener genetischer Anfangsausstattungen eine ungleiche Verteilung von Pflegeheimen oder Ähnliches zu unterschiedlichen Anteilen führen könnte.

**Kreisweite Faktoren** Hier sind hauptsächlich Umweltbedingungen und das Gesundheitssystem zu nennen. Die Wirkungen von Umweltbelastungen auf den Gesundheitszustand sind ein wichtiges Forschungsgebiet der Epidemiologie, die eine Reihe von Methoden zu ihrer Erfassung entwickelt hat (Eis 1997). Große Probleme bestehen in der Feststellung von Expositionsraten, der Isolierung von Faktoren und der Bestimmung von Wirkungslags. So wurde der Einfluss von Luftverunreinigungen durch Schwebstäube (PM<sub>10</sub>) auf die Mortalität von Individuen für die USA erst jüngst in einer groß angelegten Studie belegt (vgl. Samet et al. (2000) und Dominici et al. (2003)). Die möglichen Wirkungsrichtungen sind jedoch weitgehend klar, so können Luft-, Boden- oder Trinkwasserverunreinigungen unmittelbar zu Beeinträchtigungen der Gesundheit von Individuen führen, dies wird auch als direkter Effekt der Umwelt bezeichnet. Ein indirekter Effekt entsteht, wenn die Individuen ihr Verhalten aufgrund von Umweltbelastungen ändern (beide Begriffe nach Wille und Ried (1996), S. 176). Das Vorzeichen des zweiten Effekts ist unbestimmt, jedoch liefern die meisten Theorien einen negativen Gesamteffekt auf die Lebenserwartung (Wille und Ried (1996), S. 177-178). Empirische Untersuchungen legen ebenfalls einen negativen Gesamteffekt der Umwelt nahe, für einen knappen Überblick vgl. Wille und Ried (1996), S. 178ff. Umwelteinflüsse erscheinen ferner aufgrund von Interaktionen mit den sozioökonomischen Variablen als Kontrollvariable relevant. So haben wohlhabendere Personen größere Möglichkeiten, den Wohnort zu wechseln. Höhere Bildung kann zu einer besseren Einschätzung des durch die Umweltbelastung entstehenden Risikos sowie geeigneten Gegensteuerungsmaßnahmen führen. Entsprechend ist ein negativer Selektionseffekt zu erwarten.

## 4.2. Ökonometrische Analyse aggregierter Daten

Die Gesundheitsinfrastruktur ist eine weitere kreisspezifische Determinante. Ein vielfältiges Angebot an medizinischen Dienstleistungen sollte sich positiv auf die Mortalität auswirken (Cischinsky 2000). Insbesondere für ältere Menschen mit eingeschränktem Bewegungsradius ist die günstige Erreichbarkeit von entsprechenden Einrichtungen von Bedeutung (Wittwer-Backofen (1999), S. 159). Ähnliches gilt auch für ein gut ausgebautes Rettungsdienstnetz. Auch in diesem Fall entstehen wieder Interaktionen mit sozioökonomischen Variablen und insbesondere mit dem Gesundheitsstatus. Es ist leicht vorstellbar, dass gut informierte Individuen sowie morbidere Personen einen relativ hohen Wert auf die Qualität der Einrichtungen des Gesundheitswesens legen. Entsprechend ist das Vorzeichen des Selektionseffektes unbestimmt.

Einen möglichen Einfluss könnte auch die Siedlungsstruktur ausüben, deren Effekte bei Wittwer-Backofen (1999), S. 140, und Cischinsky (2000), S. 66, diskutiert werden. So wird unter anderem der Stressfaktor in Städten als eine mögliche Ursache für erhöhte Mortalitätsraten genannt, dem entgegen wirkt eine bessere Gesundheitsversorgung. Wie in diesem Beispiel bleibt in allen Fällen das Vorzeichen a priori unbestimmt, die Effekte erscheinen eher schwach und dürften aufgrund von Interaktionen mit sozioökonomischen Faktoren sehr schwer zu bestimmen sein.

## 4.2. Ökonometrische Analyse aggregierter Daten

Dieser Abschnitt versucht, eine kurze Einführung in die besondere Problematik der Analyse aggregierter Daten (im folgenden auch ökologische Analyse) zu geben. Genau wie das in Abschnitt 4.3 folgende Regressionsmodell ist er im Wesentlichen technischer Natur und kann nach diesem Absatz übersprungen werden, wenn der Leser nur an den Ergebnissen der Arbeit interessiert ist. Es sei lediglich ein kleiner Hinweis zur begrenzten Reichweite einer solchen Analyse gegeben, der anhand des berühmten Beispiels von Durkheim (1897) illustriert werden soll. Durkheim untersuchte die Selbstmordraten in preußischen Provinzen und fand heraus, dass sie in denen mit einem hohen protestantischen Bevölkerungsanteil und wenigen Katholiken höher lagen als in jenen mit umgekehrten Bevölkerungsverhältnissen. Sein Schluss, dass Protestanten höhere Selbstmordraten aufweisen, ist jedoch unzulässig (Robinson 1950). Eine alternative Erklärung wäre zum Beispiel, dass Katholiken stark suizidgefährdet waren, wenn sie eine marginale Bevölkerungsgruppe darstellten. Das fundamentale Problem in der ökologischen Analyse besteht somit darin, dass die Outcomes aufgrund der Datenstruktur nicht direkt den erklärenden Variablen zugeordnet werden können (Sheppard 2003). Dies führt dazu, dass sich die hier verwendeten Modelle mit Annahmen helfen müssen, die oft recht weit hergeholt erscheinen. Dazu zählen strenge



#### 4. Theoretische Überlegungen

Linearität und eine additive Struktur, Umgang mit Messfehlern und problematische Altersstandardisierungen. Die Ergebnisse des Kapitels 5 sind also nur als grobe Anhaltspunkte für Korrelationen zu werten und taugen nicht für Kausalinterpretationen.

Ein wenig technischer ausgedrückt stellt sich das Fundamentalproblem wie folgt dar. Es sind nur marginale Verteilungen gegeben, während für valide Inferenz ohne ad-hoc Annahmen gemeinsame Verteilungen notwendig sind (Greenland 2001). So können einzelne Effekte nicht sauber getrennt werden. In Tabelle 4.1 ist ein Beispiel für die komplette Verwechslung von individuellen und kontextuellen Effekten gegeben. Das Beispiel und die Darstellung im Rest dieses Absatzes sind mit leichten Modifikationen übernommen aus Greenland (2001). Angenommen, man möchte überprüfen, ob Mortalitätsunterschiede zwischen den Regionen Stadt und Land vorliegen. Als Kontrollvariable für die Bildungsstruktur steht die Bevölkerung mit Gymnasialabschluss zur Verfügung. Die Kovariate wird nur für die Grundgesamtheit, nicht aber für die Gestorbenen beobachtet. Einheitliche Sterblichkeit würde bedeuten, dass der relative Effekt der Stadt gegenüber dem Land gleich eins ist. Die beobachteten Daten sind jedoch mit einer weiten Bandbreite von Möglichkeiten kompatibel, zwei davon sind in Tabelle 4.1 aufgeführt. Im ersten Fall hat das Abitur einen negativen Effekt, die Stadt einen positiven, im zweiten Fall gilt die Umkehrung. Anhand der gegebenen Daten kann keine Aussage über die wahren Wirkungsmechanismen getroffen werden. Eine Regression würde zu dem Ergebnis kommen, dass keine Unterschiede vorlägen: die unbedingten Sterbeziffern sind in beiden Fällen gleich, der Koeffizient der Abiturquote wäre Null. Eine direkte Interpretation dieser Regression würde zu falscher Inferenz führen, weil sie nicht beachtet, dass der Koeffizient anhand der Daten nicht identifiziert ist. Zur Identifikation würde entweder die gemeinsame Verteilung von Abitur und Sterblichkeit benötigt, oder Informationen über den Regionseffekt. Hier wird auch deutlich, dass das holistische Argument zur Verteidigung ökologischer Studien (Achen und Shively 1995), individualspezifische Information sei nicht nötig, wenn sich das Interesse auf Kontexteffekte konzentriert, nicht haltbar ist. Der Kontexteffekt scheint in dieser ökologischen Regression Null zu sein. Tatsächlich ist er mit dem individuellen Effekt perfekt negativ korreliert. Die Kovariate wird nur scheinbar konstant gehalten, da die Verbindung zur abhängigen Variable gekappt ist. Die folgenden Abschnitte formalisieren einige Probleme, die bei der Analyse aggregierter Daten entstehen, und zeigen Ansätze von Lösungsmöglichkeiten auf.

**Formale Darstellung der ökologischen Analyse** Die Problematik der Analyse aggregierter Daten wird in diesem Abschnitt anhand des Querschnittsmodells illustriert. Erst im folgenden Teil 4.3 werden die Notationen und Begrifflichkeiten auf Paneldaten erweitert, um an dieser Stelle die Konzentration auf die ökologische Analyse zu ermöglichen.

## 4.2. Ökonometrische Analyse aggregierter Daten

Tabelle 4.1.: Beispiel für die Verwechslung von individuellen und kontextuellen Effekten

	<b>Stadt</b>			<b>Land</b>		
	Abitur	Abitur	Gesamt	Abitur	Abitur	Gesamt
	ja	nein		ja	nein	
<b>Daten</b>						
Gestorbene	?	?	560	?	?	560
Personenjahre (Tausend)	60	40	100	40	60	100
Rate	?	?	5,6	?	?	5,6
<b>Möglichkeit 1:</b>						
$RE_{Abi} = 2; RE_{S,L} = 7/8$						
Gestorbene	420	140	560	320	240	560
Personenjahre (Tausend)	60	40	100	40	60	100
Rate	7	3,5	5,6	8	4	5,6
<b>Möglichkeit 2:</b>						
$RE_{Abi} = 1/2; RE_{S,L} = 8/7$						
Gestorbene	240	320	560	140	420	560
Personenjahre (Tausend)	60	40	100	40	60	100
Rate	4	8	5,6	3,5	7	5,6

$RE_{Abi}$  und  $RE_{S,L}$  bezeichnen die jeweiligen relativen Effekte des Abiturs und der Siedlungsstruktur.

Quelle: Greenland (2001).

#### 4. Theoretische Überlegungen

Das individuelle Modell (4.1) lässt sich im Aggregat schreiben als:

$$(4.2) \quad \bar{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} f(\alpha + X_{ik}\beta + Z_k\gamma) .$$

$N_k$  bezeichnet die Bevölkerung des Kreises  $k$ . Für  $\beta \neq 0$  hängt  $\bar{\mu}_k$  von der innerkreislichen Verteilung von  $X$ , den individuenspezifischen Effekten  $\beta$  und den kontextuellen Effekten  $\gamma$  ab (Greenland 2002). An dieser Stelle sei zur Vereinfachung angenommen, dass  $Z_k = \bar{X}_k \forall k$ . Dies gilt in der vorliegenden Problemstellung zum Beispiel, falls das Einkommen der einzige Regressor ist.  $\beta$  wäre dann der Effekt einer Erhöhung des individuellen Einkommens auf die Sterblichkeit,  $\gamma$  derjenige einer Erhöhung des durchschnittlichen Kreiseinkommens, gegeben das individuelle Einkommen ist konstant.

Ein naives ökologisches Regressionsmodell in einer generalisierten linearen Spezifikation wäre (Greenland 2002):

$$(4.3) \quad E_k[\bar{Y}] = \bar{\mu}_k = g(a + \bar{X}_k b) .$$

Für den Spezialfall eines individuellen Modells, welches linear in den Parametern und Variablen ist, gilt:

$$(4.4) \quad \mu_{ik} = \alpha + X_{ik}\beta + \bar{X}_k\gamma$$

und

$$(4.5) \quad \bar{\mu}_k = \alpha + \bar{X}_k(\beta + \gamma) .$$

Das im letzten Abschnitt angesprochene Identifikationsproblem liegt an ausgelassenen Variablen. Wenn im linearen Modell (4.5) allgemeine Gleichgewichtseffekte oder andere mit den individuenspezifischen Regressoren korrelierte kontextuelle Variable vorhanden sind, so ist der aus (4.3) resultierende Schätzer für den individuellen Effekt verzerrt und inkonsistent. Das Problem verschlimmert sich, falls Gleichung (4.5) keine hinreichende Approximation für das wahre Modell darstellt. In diesem Fall ist die gesamte Verteilung von  $X$  in den Kreisen vonnöten, um den wahren Effekt zu identifizieren.

**Identifikation und ausgelassene Variable** Gleichung (4.5) verdeutlicht, dass in einer Analyse mit aggregierten Daten der individuelle Effekt  $\beta$  und der kontextuelle Effekt  $\gamma$  nicht separat identifiziert sind (Firebaugh 1978). Die zur Identifikation oft getroffene Annahme  $\gamma = 0$  wäre in der volkswirtschaftlichen Terminologie die Abwesenheit allgemeiner Gleichgewichtseffekte (Dow, Gonzalez, und Rosero-Bixby 2003). In dem Einkommensbeispiel würde dies bedeuten, dass das durchschnittliche Kreiseinkommen keinerlei Einfluss auf

## 4.2. Ökonometrische Analyse aggregierter Daten

die Mortalität der Individuen hat. Wie Dow, Gonzalez, und Rosero-Bixby (2003) anmerken, ist für viele Politikinterventionen jedoch der gemeinsame Effekt  $\beta + \gamma$  von Interesse, so dass eine ökologische Regression die gewünschten Ergebnisse liefern kann. Dies gilt jedoch nur bei individueller Linearität in den Variablen und  $g(v) = v$  in Gleichung (4.3).

Dow, Gonzalez, und Rosero-Bixby (2003) zeigen für das einfache lineare Modell sehr schön, welchen Einfluss ausgelassene Variable ausüben. Ausführlichere Diskussionen für nichtlineare Modelle finden sich zum Beispiel bei Sheppard (2003) sowie Neuhaus und Kalbfleisch (1998), diese werden hier jedoch nicht näher behandelt. Entscheidend ist, dass der Schätzer für den gemeinsamen Effekt  $\beta + \gamma$  fast ausschließlich auf der between-Variation (d. h. derjenigen zwischen den Kreisen) beruht (Sheppard 2003), während in einem individuellen Modell die within-Variation ein höheres Gewicht hat. Entsprechend fallen ausgelassene Variable, die auf verschiedenen Ebenen variieren, unterschiedlich stark ins Gewicht. Zur Illustration seien die beiden Spezialfälle aus Dow, Gonzalez, und Rosero-Bixby (2003) aufgeführt.

Falls die ausgelassene Variable ausschließlich zwischen den Kreisen variiert, so verschlimmern sich die Verzerrungen in der ökologischen Analyse relativ zur Arbeit mit Individualdaten. Allgemein ausgedrückt gilt dies für den Fall, in dem die between-Variation einen hohen Anteil an der Gesamtvariation der ausgelassenen Variable ausmacht. Im Rahmen der regionalen Sterblichkeitsanalyse gilt dies zum Beispiel für Umweltbelastungen oder Siedlungsstrukturen.  $R_k$  bezeichne diese ausgelassene Variable,  $\delta$  ihren wahren Effekt. Zur Vereinfachung gebe es nur einen weiteren Regressor  $X_{ik}$ , welcher mit  $R_k$  korreliert sei. Dann ergibt sich asymptotisch für den Schätzer aus dem individuellen Modell (4.4):

$$\hat{\beta}_{ind} - \beta = \delta \cdot \frac{cov[X_{ik}, R_k]}{var[X_{ik}]} = \delta \cdot \frac{cov[\bar{X}_k, R_k]}{var[X_{ik}]} .$$

Für die ökologische Regression aus (4.5) folgt asymptotisch:

$$\hat{\beta}_{agg} - \beta = \delta \cdot \frac{cov[\bar{X}_k, R_k]}{var[\bar{X}_k]} .$$

Die Differenz beider Gleichungen entsteht im Nenner, da die Varianz von  $X_{ik}$  größer ist als diejenige von  $\bar{X}_k$ . Die Verzerrung von  $\hat{\beta}_{ind}$  relativ zu  $\hat{\beta}_{agg}$  fällt umso geringer aus, je größer der Anteil der within-Variation des Einkommens ist.

Der zweite Spezialfall ergibt sich, wenn die ausgelassene Variable auf der individuellen Ebene variiert, über die Kreise hinweg jedoch homogen ist. Dies kann approximativ für die genetischen Dispositionen angenommen werden (vgl. Wittwer-Backofen (1999), S. 243-244). Formal gilt

$$cov[X_{ik}, R_{ik}] \neq 0 \quad \text{und} \quad \bar{R}_k = \bar{R} \quad \forall k .$$

#### 4. Theoretische Überlegungen

Es folgt

$$\text{cov}[\bar{X}_k, \bar{R}] = 0 \quad \text{und somit} \quad \hat{\beta}_{agg} = \beta \neq \hat{\beta}_{ind}.$$

Bei dieser Konstellation ergibt sich, dass der aggregierte Schätzer konsistent ist, während der individuelle asymptotisch verzerrt ist. Der Extremfall illustriert den allgemeineren, dass der ökologische Schätzer weniger verzerrt ist als der individuelle, wenn die within-Variation einer ausgelassenen Variable einen relativ großen Anteil an ihrer Gesamtvariation hat (Dow, Gonzalez, und Rosero-Bixby 2003).

**Nichtlinearitäten im individuellen Modell** Das ökologische Regressionsmodell (4.3) ist mit dem aggregierten Modell (4.2) nur dann identisch, wenn letzteres linear in den Parametern und Variablen ist. Mit anderen Worten, weder die Regressoren noch die abhängige Variable dürfen im individuellen Modell einer Transformation unterliegen. Dies gilt auch für Interaktionen zwischen den Regressoren (Greenland und Morgenstern 1989). Angenommen, das Einkommen sei im strukturellen Modell die einzige erklärende Variable mit positiver Wirkung auf die Mortalität, allgemeine Gleichgewichtseffekte sind nicht vorhanden. Der Effekt sei jedoch nicht linear, sondern nehme mit steigendem Einkommen ab, so dass ein quadratischer Term im Modell enthalten ist:

$$\mu_{ik} = \alpha + X_{ik}\beta_1 + X_{ik}^2\beta_2$$

Das aggregierte Modell ergibt sich als

$$\bar{\mu}_k = \alpha + \bar{X}_k\beta_1 + \frac{1}{N_k} \sum_{i=1}^{N_k} X_{ik}^2\beta_2.$$

Wegen

$$\frac{1}{N_k} \sum_{i=1}^{N_k} X_{ik}^2 \neq \bar{X}_k^2$$

lässt sich das Modell ohne das zweite Moment der kreisinternen Verteilung von  $X$  nicht schätzen. Für Interaktionen zwischen zwei Regressoren wird sogar die entsprechende gemeinsame Verteilung benötigt. Einen genaueren Überblick der Effekte von Nichtlinearitäten liefern Greenland und Robins (1994), eine Anwendung im demografischen Bereich findet sich bei Vaupel, Manton, und Stallard (1979).

**Lösungsmöglichkeiten** Wenn theoretisch relevante Variable ausgelassen werden müssen, sollte man sich zunächst die potentielle Korrelationsstruktur verdeutlichen. Wie oben dargestellt, kann es sogar sein, dass der Aggregationsprozess die Verzerrung abmildert. Scheint

das Gegenteil der Fall zu sein, sollten die Ergebnisse nach Möglichkeit mit Individualdaten verifiziert werden.

Um die aufgrund von Nichtlinearitäten entstehenden Probleme in den Griff zu bekommen, sollte möglichst viel Information über die within-Verteilung der Kovariaten herangezogen werden. Die einfachsten Möglichkeiten bestehen in der mehrdimensionalen Erfassung von Kovariaten und in der Aufnahme höherer Momente. Ein vielversprechender und genauerer Ansatz ist die Mehrebenenanalyse.

Falls eine Kovariate mehrdimensional ist, so sollten möglichst alle Dimensionen in die Regressionsgleichung aufgenommen werden (vgl. Greenland und Robins (1994) oder Guthrie und Sheppard (2001)). Rauchen als erklärende Variable beispielsweise variiert innerkreislich hinsichtlich der Zahl der Raucher und der Intensität des Tabakkonsums. Zusätzlich zu üblichen im Aggregat gemessenen Variablen wie Zigarettenverkäufen pro Kreis wäre der Prozentsatz der Raucher als Regressor wünschenswert (Greenland und Robins 1994). Bei eindimensionalen Regressoren können die Probleme auch durch eine Taylor-Approximation höherer Ordnung abgemildert werden (vgl. Lasserre, Guihenneuc-Jouyaux, und Richardson (2000) oder Guthrie und Sheppard (2001)). In der Regel sollte eine Entwicklung zweiter Ordnung genügen (Plummer und Clayton 1996), so dass die Aufnahme der Varianz des Regressors in die Schätzgleichung dieses Ziel erreicht. Unter zusätzlichen Annahmen (log-lineares Modell,  $X$  multivariat normalverteilt mit bekannter Varianz) beseitigt der zusätzliche Term die Verzerrung (Richardson, Stucker, und Hemon 1987).

Die Nützlichkeit dieser zusätzlichen Informationen über die Verteilung von  $X$  hängt sehr von der Anwendung ab, für wenig sinnvolle Ergebnisse vgl. zum Beispiel die Simulationen in Guthrie und Sheppard (2001), S. 147. Der im Kontext dieser Arbeit vermutlich optimale Ansatz wurde von Prentice und Sheppard (1995) entwickelt. Die Autoren verwenden ein log-lineares Modell der Form

$$\bar{\mu}_k = \eta_k \exp(\alpha + X_{ik}\beta + Z_k\gamma)$$

und setzen voraus, dass für die individuenspezifischen Variable  $X_{ik}$  eine Stichprobe aus jedem Kreis vorliegt. Die Identifikation wird durch Annahmen bezüglich der regionspezifischen Baseline-Raten  $\eta_k$  erreicht, ohne dass Individualdaten zur abhängigen Variable vorliegen (Greenland 2001). Der Ansatz ist besonders geeignet, wenn die abhängige Variable sehr kleine Werte annimmt, wie zum Beispiel Erkrankungs- oder Sterberaten. Guthrie und Sheppard (2001) zeigen in Simulationsstudien, dass diese Form der Mehrebenenanalyse selbst bei relativ kleinen Stichproben gute Ergebnisse liefert.

#### 4. Theoretische Überlegungen

**Nicht vergleichbare Standardisierung** Rosenbaum und Rubin (1984) zeigen, dass es zu erheblichen Verzerrungen kommen kann, wenn erklärende und abhängige Variable auf verschiedene Weise standardisiert sind. Dies geschähe zum Beispiel mit Alter und Einkommen, wenn standardisierte Sterbeziffern oder Lebenserwartungen als abhängige Variable gewählt würden. Die Daten zum Einkommen stehen nicht altersabhängig zur Verfügung (s. Abschnitt 5.1), so dass sie unstandardisiert verwendet werden müssten. Das Argument von Rosenbaum und Rubin (1984) ist eine einfache Anwendung des Frisch-Waugh-Theorems. Wenn Sterblichkeit und Einkommen vom Alter abhängen, so können beide Variable durch Regressionen auf das Alter standardisiert werden. Eine Regression der Residuen aufeinander liefert unverzerrte Koeffizienten. Wird nur die Sterblichkeit in irgendeiner Form standardisiert, so ist der Schätzer für den Einfluss des Einkommens verzerrt. Für die hier gegebene Datenstruktur schlagen Rosenbaum und Rubin (1984) vor, rohe Sterbeziffern als abhängige Variable zu verwenden und die ersten  $J$  Momente der Altersverteilung als Regressoren aufzunehmen. Dieser Weg wird auch in dieser Arbeit besprochen, wobei sich Werte für  $J$  von drei oder vier als sinnvoll erwiesen.

### 4.3. Ein einfaches Regressionsmodell

Die meisten der in dieser Arbeit verwendeten Daten liegen im Panelformat vor, wobei der Kreis die Beobachtungseinheit bildet. Die unterschiedliche Bedeutung der Termini between- und within-Variation in den Zusammenhängen von Paneldaten auf Individualebene, aggregierten Querschnittsdaten und ökologischen Paneldaten kann zu einigen Verwirrungen führen. Tabelle 4.2 versucht, ein wenig Klarheit in die Begrifflichkeit zu bringen. Im folgenden wird ein ökologisches Panelmodell aus einem individuellen Panel- und ökologisches Querschnittsmodell hergeleitet. Grundvoraussetzung ist dabei die Linearitätsannahme, die im vorhergehenden Abschnitt ausführlich kritisiert wurde. Im Rahmen der vorliegenden Arbeit kann sie jedoch nur dahingehend aufgeweicht werden, dass einige höhere Momente von Regressoren in die Schätzgleichung aufgenommen werden. Erweiterungen mit allgemeineren Spezifikationen müssen zukünftigen Projekten auf breiterer Datenbasis vorbehalten bleiben.

Ein lineares Panelmodell auf individueller Ebene mit unbeobachteten Effekten (Wooldridge (2002), S. 251) ist gegeben als:

$$(4.6) \quad Y_{it} = \alpha + X_{it}\beta + c_i + u_{it} .$$

In der vorliegenden Anwendung bezeichnet  $Y_{it}$  eine Sterbewahrscheinlichkeit, somit stellt (4.6) ein Lineares Wahrscheinlichkeitsmodell dar. Ein Querschnittsmodell mit unbeobach-

Tabelle 4.2.: Between- und within-Variation in Paneldaten und ökologischen Regressionen

Datentyp		between-Variation	within-Variation
<b>individuelles Panel</b>	Variation	zwischen Individuen	über die Zeit
	Beobachtungseinheit	individuelle Mittelwerte (über die Zeit)	einzelne Individuen
<b>ökologischer Querschnitt</b>	Variation	zwischen Kreisen	zwischen Individuen
	Beobachtungseinheit	kreisspezifische Mittelwerte (über Individuen)	einzelne Individuen (innerhalb eines Kreises)
<b>ökologisches Panel</b>	Variation	zwischen Kreisen	über die Zeit
	Beobachtungseinheit	kreisspezifische Mittelwerte (über Individuen und die Zeit)	kreisspezifische Mittelwerte (über Individuen)

teten Effekten auf Kreisebene lässt sich schreiben als:

$$(4.7) \quad Y_{ik} = \alpha + X_{ik}\beta + d_k + u_{ik} .$$

Zusammen ergeben die beiden Gleichungen ein Paneldatenmodell, welches Clustereffekte berücksichtigt:

$$(4.8) \quad Y_{itk} = \alpha + X_{itk}\beta + c_{ik} + d_{tk} + u_{itk} .$$

Der individuenspezifische Effekt  $c_{ik}$  kann nun in zwei Komponenten partitioniert werden  $c_{ik} = c_{1ik} + c_{2i}$ .  $c_{1ik}$  hängt sowohl vom Kreis als auch vom Individuum ab,  $c_{2i}$  ist konstant über die Kreise. Analog kann man den kreisspezifischen Effekt in eine zeitabhängige und eine zeitkonstante Komponente aufteilen:  $d_{tk} = d_{1tk} + d_{2k}$ . Beide Partitionierungen eingesetzt in (4.8) und über die Individuen aggregiert ergibt:

$$(4.9) \quad \bar{Y}_{tk} = \alpha + \bar{X}_{tk}\beta + \bar{c}_{1k} + \bar{c}_2 + d_{1tk} + d_{2k} + \bar{u}_{tk} .$$

Mit  $\psi = \alpha + \bar{c}_2$ ,  $\lambda_k = \bar{c}_{1k} + d_{2k}$  und  $\varepsilon_{tk} = \bar{u}_{tk} + d_{1tk}$  lässt sich dies schreiben als:

$$(4.10) \quad \bar{Y}_{tk} = \psi + \bar{X}_{tk}\beta + \lambda_k + \varepsilon_{tk} .$$

Die abhängige Variable ist die rohe Sterberate in der Kreisbevölkerung,  $\bar{X}_{tk}$  kann Momente von individuellen Regressoren und kreisspezifische Effekte enthalten. Beide können



#### 4. Theoretische Überlegungen

Tabelle 4.3.: Beispiele für die unbeobachteten Effekte, welche für die Annahmen von Random Effects relevant sind

Formales Symbol	Art des unbeobachteter Faktors	Beispiele
$\bar{u}_{tk}$	Individuell, zeitabhängig	Stress am Arbeitsplatz, Freizeitaktivitäten
$\bar{c}_{1k}$	Individuell, zeitkonstant, variiert auf Kreisebene	Ernährungsgewohnheiten, (approximativ zeitkonstant)
$d_{1tk}$	Kreisweit, zeitabhängig	Trinkwasserqualität, Rettungsdienstversorgung, Kriminalitätsrate
$d_{2k}$	Kreisweit, zeitkonstant	sozialer Kontext, klimatische Grundbedingungen

zeitabhängig sein. Aus Gleichung (4.9) wird erneut deutlich, dass solche Effekte, die individuell variieren, aber zwischen den Kreisen konstant sind, die aggregierten Schätzkoeffizienten nicht beeinflussen:  $\bar{c}_2$  wird Teil der Konstante.

Zur Schätzung von (4.10) bietet sich ein Random Effects-Ansatz an. Die dazu notwendigen Annahmen bestehen aus (Wooldridge (2002), S. 257ff.):

$$(4.11) \quad E[\varepsilon_{tk}|X_k, \lambda_k] = 0$$

$$(4.12) \quad E[\lambda_k|X_k] = 0$$

$$(4.13) \quad \text{Rang}(E[X_k' \Omega^{-1} X_k]) = L$$

$$(4.14) \quad \Omega = \sigma_\varepsilon^2 I_T + \sigma_\lambda^2 i_T i_T' ,$$

wobei  $\Omega \equiv E[(\lambda_k + \varepsilon_k)(\lambda_k + \varepsilon_k)']$ .  $L$  sei die Anzahl der Elemente in  $X$  (zur Vereinfachung der Notation beinhalte  $X$  in diesem Fall ausnahmsweise eine Konstante) und  $T$  die Zahl der Perioden.  $I_T$  bezeichne die Identitätsmatrix der Dimension  $T$ ,  $i_T$  den Einheitsvektor mit  $T$  Elementen. Tabelle 4.3 schlüsselt die für diese Annahmen relevanten unbeobachteten Effekte näher auf.

Tabelle 4.4.: Aufschlüsselung der Regressoren

Formales Symbol	Variablentyp	Klassifizierung der vorhandenen Regressoren
$\bar{X}_{1tk}$	Individuell, zeitabhängig	Altersverteilung, Einkommen, Bildung, Wanderung, Ausländer, Arbeitslosenunterstützungsempfänger, Sozialhilfeempfänger
$X_{2tk}$	Kreisweit, zeitabhängig	Luftbelastung, Bettenversorgung Flächennutzung,
$X_{3k}$	Kreisweit, zeitkonstant	Dummy für die Stadtkreise

Analog zu der Partitionierung der unbeobachteten Effekte lassen sich auch die vorhandenen Regressoren aufteilen. Dies geschieht in Tabelle 4.4. Die kritischen Annahmen für den Random Effects-Schätzer sind (4.11) und (4.12). Bei der ersten Annahme erscheint insbesondere die strikte Exogenität der Regressoren restriktiv. So darf beispielsweise eine Veränderung des Einkommens in Periode 1 keinen Einfluss auf die sterblichkeitsrelevanten Freizeitaktivitäten in Periode 2 haben. Dies erscheint auf individueller Ebene eher unwahrscheinlich, auf aggregiertem Niveau ist es jedoch durchaus möglich, dass sich die Effekte ausgleichen. Ferner stellt sich die Frage nach der gesuchten Kausalität: in dieser Untersuchung erscheint es angebracht, vom Einkommen als Kausaleffekt auszugehen, unabhängig davon, ob es direkt (z. B. Zugang zum Gesundheitswesen) oder indirekt (z. B. Golfspielen statt Sportklettern in der Freizeit) wirkt. Die Interpretation der geschätzten Koeffizienten sollte sich aufgrund der in Abschnitt 4.1 genannten Simultanitätsproblematik ohnehin auf eine partielle Korrelation beschränken, von Kausalität kann nicht gesprochen werden. Bei der gegebenen Datenmenge ist Annahme (4.11) jedoch nahezu unverzichtbar, da die Zahl der Beobachtungen für dynamische Paneldatenmodelle nicht ausreicht.

Annahme (4.12) wird in der Paneldatenliteratur ebenfalls sehr kritisch beleuchtet (Wooldridge (2002), S.265). Im gegebenen Kontext erscheint sie aufgrund der Durchschnittsbildung jedoch nicht so restriktiv wie im Falle von Individualdaten. Der problematischste Fall im Rahmen der in Tabelle 4.3 genannten Beispiele ist die potenzielle Korrelation von Einkommen und Bildung sowie sozialem Kontext. Congdon (1995) findet in

#### 4. Theoretische Überlegungen

einer Analyse der kleinräumigen Mortalität in London Effekte des sozialen Kontextes auch wenn sozioökonomische Charakteristika konstant gehalten werden. Da die Beobachtungseinheiten hier deutlich größer sind, bleibt zu hoffen, dass durch die Aggregation heterogener Distrikte und die Verwendung der Stadtkreisdummy die Annahme approximativ erfüllt ist. Alternativ wäre eine Fixed Effects-Schätzung möglich. Diese ist jedoch für die Fragestellung dieser Arbeit nicht geeignet, da sie lediglich die Variation innerhalb der Kreise über die Zeit verwendet. Von Interesse im Sinne einer Querschnittsbetrachtung sind hier jedoch die Unterschiede zwischen den Kreisen.

Die Rangbedingung (4.13) dient lediglich der Identifikation und ist in aller Regel erfüllt (Wooldridge (2002), S. 154). Die Homoskedastieannahme (4.14) ist notwendig für statistische Inferenz auf Basis der verwendeten Modelle. Auch sie erscheint im Kontext von aggregierten Daten weniger restriktiv als auf individueller Ebene, kann jedoch für ein lineares Wahrscheinlichkeitsmodell per definitionem nicht erfüllt sein. Dies kann lediglich approximativ der Fall sein.

Als Essenz dieses Abschnitts bleibt festzuhalten, dass für die Analyse der vorliegenden Daten eine Menge Annahmen notwendig sind, die wenig plausibel erscheinen. Dies trifft insbesondere auf die strenge Linearität und Additivität des Modells zu. Einige der üblichen Probleme der ökologischen Analyse konnten auf theoretischer Ebene abgemildert werden, so zum Beispiel der verstärkte Einfluss von Messfehlern. Mit diesen Ergebnissen und der damit beschränkten Interpretationsreichweite im Hinterkopf wendet sich die Arbeit nun der konkreten Datenauswertung zu.

## 5. Regressionsanalyse mit Paneldaten

In diesem Kapitel wird versucht, den Einfluss der in Abschnitt 4.1 dargelegten Faktoren zu überprüfen. Zunächst werden die hier verwendeten Daten des Statistischen Landesamtes Baden-Württemberg, welche in Panelform vorliegen, näher beschrieben. Im Zuge dessen werden die Hypothesen bezüglich der einzelnen Regressoren formuliert. Es folgt die Darstellung der Ergebnisse.

### 5.1. Datensatzbeschreibung und Hypothesenformulierung

In Tabelle 5.1 wird ein knapper Überblick über die nach Jahren vorhandenen Daten gegeben. Es wird deutlich, dass die Sterblichkeits- und Wanderungsdaten zwar in einer langen Zeitreihe vorliegen, die Mehrzahl der Regressoren jedoch nur für einige Abschnitte vorhanden sind. Insgesamt kristallisieren sich für die Regressionen drei interessante Zeitreihen heraus. Am interessantesten ist der Zeitraum 1999 bis 2002, wo die besten Daten zu sozioökonomischen Charakteristika vorliegen, insbesondere die Bildung am Wohnort und das geschlechtsspezifische Renteneinkommen. Auch die Immigrationsstatistiken gibt es nur für diesen Zeitraum. Interessant sind auch die Zeitreihen von 1983 bis 1995, wo Bildung am Arbeitsort und die Daten der Einkommensteuerstatistik vorliegen. Beide sind jedoch nicht geschlechtsspezifisch vorhanden. Abschließend zu nennen bleibt der Zeitraum von 1995 bis 2002, wo relativ viele Regressoren vorhanden sind. Größtes Manko ist hier das Nichtvorhandensein einer Einkommensgröße.

Sofern möglich wurden alle Variablen geschlechtsspezifisch und für drei verschiedene Altersgruppen berechnet: sämtliche Alter, die 18- bis 65-jährigen, sowie die Personen ab 66 Jahren. Die gemeinsamen Verteilungen lagen jedoch nicht für alle Regressoren vor, so dass die in Abschnitt 4.2 dargestellte Kritik von Rosenbaum und Rubin (1984) zutrifft. In diesen Fällen besteht jedoch ein Trade-off zwischen zwei Fehlern. Eine Stratifizierung der abhängigen Variable, ohne dasselbe für den Regressor zu tun, beinhaltet die Annahme, dass letztgenannter unkorreliert mit Alter und Geschlecht ist. Aggregiert man alle Daten und

## 5. Regressionsanalyse mit Paneldaten

hält die demografischen Charakteristika konstant, so gehen durch die Aggregation heterogener Risiken wertvolle Informationen verloren. All dies läuft auf ein Abwägen zwischen zwei Messfehlern hinaus, welches nur im Einzelfall entschieden werden kann. Die folgenden Absätze beschreiben die Daten zu den verwendeten Variablen detailliert. In Tabelle A.14 finden sich die entsprechenden deskriptiven Statistiken.

**Sterblichkeit und Altersaufbau** Aus den bereits in Kapitel 2 dargestellten Bevölkerungs- und Sterblichkeitsdaten wurden rohe Sterbeziffern sowie die ersten vier Momente der Altersverteilung berechnet. Wie oben diskutiert wurde, sind die Daten für Deutsche als hinreichend gut einzustufen. Für Personen mit ausländischer Staatsbürgerschaft scheinen die Sterbeziffern jedoch zu gering zu sein. Dies könnte zu systematischen Verzerrungen führen, welche durch die Aufnahme des Ausländeranteils als Regressor möglicherweise nur unzureichend kontrolliert werden. Daher wurden die genannten Kennziffern für die deutsche und für die Gesamtbevölkerung berechnet. Eine Analyse ausschließlich für Ausländer verbot sich aufgrund der geringen Fallzahl in vielen Kreisen.

**Einkommen** Zu den Einkommensverhältnissen der Kreisbewohner stehen Daten aus der Lohn- und Einkommensteuerstatistik der Jahre 1983, 1986, 1989, 1992 und 1995 zur Verfügung. Die Daten sind leider mit mehreren Messfehlern behaftet, welche in den nächsten Absätzen sukzessive behandelt werden. An jede Illustration schließt sich die zur Lösung der Messfehlerproblematik notwendige Annahme an.

Das erste Problem betrifft den erfassten Personenkreis. Dieser besteht aus allen Personen, die in einem nicht geringfügigen abhängigen Beschäftigungsverhältnis stehen, sowie denjenigen, die eine Einkommensteuererklärung abgegeben haben. Damit dürften insbesondere die Einkünfte von Personen am unteren Ende der Einkommenskala schlecht abgebildet sein, weil zum Beispiel Sozialhilfeempfänger in der Statistik nicht erfasst werden.

Die nicht erfassten Personen stellen einen nicht zu korrigierenden Messfehler dar. Es bleibt nur die Annahme, dass die Lücken in allen Kreisen gleichmäßig sind. Eine gewisse Milderung der Heterogenität in dieser Beziehung kann durch die Aufnahme der Quoten von Arbeitslosenunterstützungs- und Sozialhilfeempfängern in die Regression erreicht werden. Allerdings verbietet sich die Interpretation des Einkommenskoeffizienten als globale Korrelation von Einkommen und Sterblichkeit, da Nichtlinearitäten zu erwarten sind (Klein 2001) und der Bereich mit der vermutlich stärksten Korrelation (d.h. sehr geringe Einkommen) nicht erfasst ist.

Die Beobachtungseinheiten in der Einkommensteuerstatistik sind nicht uneingeschränkt miteinander vergleichbar. Zwischen Einzelpersonen, Ehepaaren, Haushalten mit Kindern

5.1. Datensatzbeschreibung und Hypothesenformulierung

Tabelle 5.1.: Übersicht der verfügbaren Daten nach Jahren.

Jahr	Lebende*	Gestorbene*	Frauenquote	Ausländerquote*	Bildung (Wohnort)*	Bildung (Arbeitsort)*	Bildung (Arbeitsort)**	Einkommen (EKSt)**	Einkommen (VDR)*	Erwerbsminderungsrenten**	Arbeitslose**	Sozialhilfeempfänger*	Stadtdummy	Wanderung*	Krankenbetten	Flächennutzung	Emissionen (SO <sub>2</sub> , CO)	Emissionen (Staub)	Immissionen	
1981	x	x	x	x									x							
1982	x	x	x	x									x							
1983	x	x	x	x	x		x				(x)		x							
1984	x	x	x	x	x	x	(x)	x			(x)		x							
1985	x	x	x	x	x	x	(x)	x			x		x							
1986	x	x	x	x	x	x	x	x			x		x							
1987	x	x	x	x	x	x	(x)	x			x		x							
1988	x	x	x	x	x	x	(x)	x			x		x							
1989	x	x	x	x	x	x	x	x			x		x							
1990	x	x	x	x	x	x	(x)	x			x		x				x			
1991	x	x	x	x	x	x	(x)	x			x		x				(x)	(x)		
1992	x	x	x	x	x	x	x	x			x		x				(x)	(x)		
1993	x	x	x	x	x	x	(x)	x			x		x				(x)	(x)		
1994	x	x	x	x	x	x	(x)	x			x		x				(x)	(x)		
1995	x	x	x	x	x	x	x	x			x		x				x			
1996	x	x	x	x	x	x					x		x				(x)			
1997	x	x	x	x	x	x					x		x				(x)			
1998	x	x	x	x	x	x					x		x				(x)			
1999	x	x	x	x	x	x					x		x				(x)			x
2000	x	x	x	x	x	x					x		x				(x)			x
2001	x	x	x	x	x	x					x		x				(x)			x
2002	x	x	x	x	x	x					x		x				(x)			x

x : Daten vorhanden, (x): Daten inter-/extrapoliert. \* Differenziert nach Geschlecht, \*\* Keine Differenzierung nach Geschlecht.  
 Quelle: Statistisches Landesamt Baden-Württemberg, VDR, Sozialministerium Baden-Württemberg.

## 5. Regressionsanalyse mit Paneldaten

etc. wird nicht unterschieden. Daher kann auch nicht berechnet werden, wie viele Personen in der Einkommensteuerstatistik erfasst sind. Die rohe Zahl der Steuerpflichtigen dividiert durch die Bevölkerungszahl ergibt Werte zwischen 0,36 und 0,47. Diese sind jedoch nicht aussagefähig, da hinter einem Steuerpflichtigen eine ganze Familie stehen kann.

Die Nichtvergleichbarkeit der Beobachtungseinheiten lässt sich durch die Annahme lösen, dass nur das Haushaltseinkommen für die Sterblichkeit relevant ist. Dafür würden Wirkungshypothesen wie diejenigen über den Lebensstil oder Ernährungsgewohnheiten sprechen, dagegen jedoch Argumente, die auf dem Zugang zum Gesundheitswesen basieren. Alternativ kann angenommen werden, dass die gemeinsame Verteilung von Haushaltsstruktur und Einkommen über die Kreise gleich ist, nachdem die anderen Regressoren konstant gehalten wurden.

Der dritte und letzte wichtige Messfehler betrifft die Einkommenshöhe. Das steuerpflichtige Jahreseinkommen steht nur in sechs (1983, 1986) bzw. elf (1989, 1992, 1995) Klassen zur Verfügung, wobei die nach oben hin offene Klasse in allen Jahren bei 100.000 DM beginnt. Ferner ist das steuerpflichtige Einkommen aufgrund von Steuerfreibeträgen, nicht angegebenen Zinserträgen im Ausland o. Ä. nicht unbedingt mit dem tatsächlich relevanten Einkommen identisch. Durch den Abzug von Kinderfreibeträgen und die Nichtberücksichtigung des Kindergeldes im steuerpflichtigen Einkommen wird die Mehrpersonenproblematik für Haushalte mit Kindern zumindestens abgemildert.

Zur Berechnung der Verteilungsmomente aus dem klassenweise erfassten Einkommen wurde eine Gleichverteilungsannahme innerhalb der Kategorien getroffen. In den mittleren Klassen erscheint diese insbesondere für die späteren Jahrgänge wenig restriktiv, da die Gruppen hinreichend klein wirken. Problematisch ist sie jedoch in den beiden extremen Kategorien. In der untersten (1 DM bis 8000 DM bzw. 5000 DM) werden mit dieser Annahme zu viele extrem geringe Einkommen erzeugt. Sie lässt sich allerdings dadurch rechtfertigen, dass viele Einkommen in dieser Klasse überhaupt nicht erfasst sind. Wenn die Zahl der nicht berücksichtigten Einkommen mit den vorhandenen korreliert ist, so ist die Verwendung von zu geringen Werten ein Proxy für eine zu geringe Zahl solcher Einkommen. In der höchsten Einkommensklasse mit der unteren Grenze 100.000 DM wurden drei verschiedene obere Beschränkungen gewählt: 125.000, 150.000 und 200.000 DM. Dies erscheint zwar als ungenügende Approximation, jedoch sprechen zwei pragmatische Gründe gegen genauere Verfahren. Einerseits legen frühere Forschungsergebnisse nahe (Klein 2001), dass der Einkommenseffekt insbesondere im niedrigen Bereich groß ist. In diesem Fall spricht Einiges dafür, den Einfluss von extremen Ausprägungen auf Mittelwert und Varianz zu begrenzen. Dagegen würde für das zweite Moment lediglich die Hypothese sprechen, dass Ungleichheit zu höherer Sterblichkeit in den unteren Einkommensklassen führt (Wilkinson 1996). Aller-

### 5.1. Datensatzbeschreibung und Hypothesenformulierung

dings ist unklar, was die Referenzebene ist (Smith (1999), S. 165). Intuitiv erscheinen die unmittelbare Nachbarschaft oder die gesamte deutsche Gesellschaft plausibler als die Kreisebene. Nach der Logik der in Abschnitt 4.2 aufgeführten Diskussion würde der Fehler dann durch die kreisweite Aggregation abgemildert oder wäre für innerdeutsche Interpretationen nicht vorhanden. Schließlich sind die Einkommensdaten mit so vielen Fehlern behaftet, dass der Einsatz von aufwändigen Verfahren zur Korrektur dieser vergleichsweise kleinen Ungleichheit nicht gerechtfertigt erscheint. Sinnvoller wirkt eine Konzentration auf andere Datenquellen, wie die unten beschriebenen Daten vom VDR oder aus stichprobenmäßigen Erhebungen wie der Einkommens- und Verbrauchsstichprobe oder dem Sozioökonomischen Panel.

In den Regressionen werden Mittelwerte und Varianzen auf Basis der Preise von 1981 verwendet, die mit dem Verbraucherpreisindex bzw. dem Preisindex für die Lebenshaltung aller privaten Haushalte errechnet wurden (Statistisches Bundesamt 2003). Die Daten wurden auf monatliche Einkommen in Euro umgerechnet, damit sie in derselben Einheit wie die Renteneinkommen (s. nächster Absatz) vorliegen. Ferner wurden die Werte mit einer einfachen arithmetischen Mittelung interpoliert, so dass eine geschlossene Zeitreihe von 1983 bis 1995 vorliegt. Diese Maßnahme erscheint nicht besonders restriktiv, da das Einkommen auf Kreisebene geringen Zufallsschwankungen unterliegen sollte. Alternative Spezifikationen nur mit vorhandenen Werten sowie Dreijahresmitteln der restlichen Variablen lieferten gleichwertige Ergebnisse. Die Hypothese bezüglich der Einkünfte ist klar: Sie sollten eine negative Korrelation mit der Sterblichkeit aufweisen. Sowohl die vermuteten Nichtlinearitäten im unteren Bereich als auch die Ungleichheitshypothese deuten auf eine positive Korrelation mit der Varianz des Einkommens hin.

**Renteneinkommen** Vom Verband der Deutschen Rentenversicherungsträger wurde für das Jahr 2002 die Zahl der Rentner und das durchschnittliche Renteneinkommen differenziert nach Alter und Geschlecht bereitgestellt.

Zum einen wurde für jeden Kreis das durchschnittliche Einkommen aus Altersrenten für Personen über 65 Jahren berechnet. Dies ignoriert zwar viele interessante Aspekte der Daten ebenso wie Selektionseffekte, sollte aber eine wichtige Messgröße der Einkommenssituation der Rentnergeneration bieten, da sie auf deren Lebenseinkommen beruht. Falls keine Kohorteneffekte vorliegen, so können die Daten auch für vorhergehende Jahre verwendet werden und bieten zudem einen Anhaltspunkt für die Einkünfte der sozialversicherungspflichtig Beschäftigten. Wenn die Annahme nur näherungsweise gilt, so entsteht ein Messfehler. Die Werte für 2002 wurden für die Jahre 1999 bis 2001 unverändert über-



## 5. Regressionsanalyse mit Paneldaten

nommen. Die Hypothese ist wiederum eindeutig: Durchschnittliche Renteneinkommen und Sterblichkeit sollten eine negative Korrelation aufweisen.

Weiterhin wurde aus den VDR-Daten die Zahl der Empfänger von Erwerbsminderungsrenten im Alter von 18 bis 65 Jahren berechnet, für die Jahre 1999 bis 2001 übernommen und als Quote an der Gesamtzahl der Bevölkerung diesen Alters in die Regression aufgenommen. Ein hoher Anteil der Bezieher von Erwerbsminderungsrenten würde auf eine Überrepräsentation von gesundheitsschädigenden Berufen hindeuten, das Vorzeichen sollte entsprechend positiv sein.

**Bildung** Die Daten zur Bildung entstammen der Beschäftigtenstatistik und beinhalten drei Kategorien: Personen ohne abgeschlossene Berufsausbildung, mit abgeschlossener Berufsausbildung und solche mit Universitäts-, Fachhochschul-, oder höherem Fachschulabschluss. Damit wird nur die Bildung der sozialversicherungspflichtig Beschäftigten erfasst und diese nur zu etwa 90%, für den Rest liegt in der Beschäftigtenstatistik keine Bildungsangabe vor. Die in dieser Bevölkerungsgruppe gemessene Verteilung der Bildungsabschlüsse muss für die Gesamtbevölkerung gelten, damit individuelle Effekte der Ausbildung identifiziert werden können. Der Fall ist mit Sicherheit nicht gegeben, da Akademiker und Personen mit abgeschlossener Berufsausbildung in der Gruppe der Selbständigen und Freiberufler überrepräsentiert sind, gegenteiliges gilt allerdings für Arbeitslose und Sozialhilfeempfänger. Ferner ist der Anteil der Personen mit höheren Bildungsabschlüssen im Zeitablauf gestiegen, so dass sich die Werte nicht ohne weiteres auf die Rentnergeneration übertragen lassen. Allgemein gesprochen ist ein erheblicher Messfehler in den Daten vorhanden. Unter der Annahme, dass dieser sich gleichmäßig über die Kreise verteilt, sind jedoch qualitative Aussagen möglich. Daten zum Bildungsabschluss der Arbeitnehmer am Wohnort differenziert nach Geschlecht und Staatsangehörigkeit sind nur für die Jahre 1999 bis 2002 vorhanden. Längere Zeitreihen existieren nur für die Bildungsindikatoren am Arbeitsort. Von 1983 an liegen nach Staatsangehörigkeit differenzierte Daten vor, ab 1995 ist auch eine Unterscheidung nach Geschlecht möglich. Ausgehend davon wurden für die beiden Arten von Abschlüssen Quoten der Gesamtbeschäftigtenzahl berechnet, die Arbeitnehmer ohne Bildungsabschluss bilden die ausgelassene Kategorie.

Die Ausprägungen der Indikatoren sind zum Teil abhängig davon, ob sie am Wohnort oder am Arbeitsort gemessen sind. In dieser Arbeit ist nur die Bildung am Wohnort relevant. Für den Zeitraum von 1999 bis 2002 wurden daher vergleichende Regressionen mit allen drei Variablen durchgeführt. Alle Spezifikationen lieferten ähnliche Resultate. Die Vorzeichen der Koeffizienten unterschieden sich nur in einer sehr geringen Zahl von Fällen, Signifikanzniveaus und Größenordnungen waren weitgehend vergleichbar. Alle drei Krite-

### 5.1. Datensatzbeschreibung und Hypothesenformulierung

rien hingen weit mehr von den zusätzlichen Kovariaten ab als von der Spezifikation der Bildungsvariable. Nichtsdestotrotz sollte bei der Verwendung der Bildungsindikatoren am Arbeitsort berücksichtigt werden, dass sie nicht die gewünschte Variable konstant halten. In weiter zurückliegenden Jahren muss die Äquivalenz der Ergebnisse nicht zwangsläufig halten.

Da die niedrigste Bildungskategorie ausgelassen ist, sollten beide in den Regressionen verwendeten Quoten ein negatives Vorzeichen aufweisen. Der Koeffizient für den Anteil der Personen mit Universitäts-, Fachhochschul- oder höherem Fachschulabschluss sollte im Betrag größer sein.

**Erwerbstätige** Aus den soeben beschriebenen Daten der Beschäftigtenstatistik wurde zusätzlich die Quote der abhängig Beschäftigten an der Bevölkerung zwischen 18 und 65 Jahren gewonnen. Diese bietet zumindestens den Ansatz einer Berücksichtigung der Tatsache, dass die für die Arbeitnehmer gemessenen Bildungscharakteristika nicht für die restliche Bevölkerung gelten. Allerdings bildet die Erwerbstätigenquote nicht nur die Bildung ab: Allgemein stellt sie die durchschnittlichen Charakteristika der sozialversicherungspflichtig Beschäftigten im Vergleich zur restlichen Bevölkerung im entsprechenden Alter dar. Da letztere aus höchst unterschiedlichen Personen besteht, ist das Vorzeichen a priori unklar. Während die Mehrzahl der Selbständigen, Freiberufler und Beamten bessere Risiken darstellen dürfte, ist dies für die Beschäftigten ohne Bildungsangabe und Sozialleistungsempfänger vermutlich umgekehrt. Die letztgenannten können jedoch berücksichtigt werden. A priori unklar ist die Sterblichkeit der Hausfrauen und -männer relativ zu den abhängig Beschäftigten.

**Arbeitslosenunterstützung und Sozialhilfe** Daten zu den Arbeitslosen liegen in einer geschlossenen Zeitreihe von 1985 bis 2002 vor, es wurden Stichtagsdaten jeweils zum 30.6. verwendet. Die Daten des ersten vorhandenen Jahres wurden in die Jahre 1983 und 1984 fortgeschrieben, um die Zeitreihe der Bildungsvariablen voll ausnutzen zu können. Dies erscheint gerechtfertigt, da kurzfristige Schwankungen der Arbeitslosigkeit aus theoretischer Sicht keinen starken Effekt auf die Sterblichkeit haben sollten. Wie Proberechnungen zeigten, änderten sich Koeffizienten und Signifikanz der Arbeitslosigkeit nicht durch die Extrapolation im Vergleich zu kürzeren Zeitreihen. Als Regressor verwendet wurde der Anteil der Arbeitslosen an der Bevölkerung zwischen 18 und 65 Jahren, a priori ist eine positive Korrelation mit der Sterblichkeit zu erwarten.

Die Empfänger laufender Hilfe zum Lebensunterhalt liegen für die Jahre 1995 bis 2002 jeweils zum Stichtag 1.1. vor. Eingang als Regressoren finden deren Quoten bezogen auf

## 5. Regressionsanalyse mit Paneldaten

die Gesamtzahl der Frauen bzw. Männer, zu erwarten ist ein positiver Zusammenhang mit der Sterblichkeit.

**Wanderung** Die Wanderungsdaten wurden bereits oben in Kapitel 2 beschrieben und für die Regressionen lediglich in Quoten umgewandelt, wobei für die drei Altersgruppen jeweils geschlechtsspezifische Werte berechnet werden konnten. Mehrere Spezifikationen wurden berechnet, Verwendung fanden nach umfangreichen Proberechnungen letztlich die Quoten der Zu- und Abwanderung. Nur die Gesamtwanderung oder Salden zu verwenden erschien zu undifferenziert. Die Aufnahme der Herkünfte bzw. Ziele (Deutschland/Ausland) in die Zu- und Abwanderungen lieferte keine zusätzlichen Erkenntnisse, sondern lediglich Multikollinearitätsprobleme. Letzteres spricht gegen die Existenz eines Messfehlers der in Kapitel 3 simulierten Form (vgl. S. 18 ff.). Wären unberücksichtigte Abwanderungen ins Ausland mit den erfassten korreliert, so sollte der entsprechende Koeffizient aufgrund der Überschätzung des Bestandes ein negatives Vorzeichen aufweisen, wobei der Effekt unabhängig von den anderen Wanderungsbewegungen sein sollte.

A priori ist während der Berufstätigkeit aufgrund des guten Gesundheitszustandes der mobilen Personen ein negatives Vorzeichen für die Zuwanderungsquote zu erwarten, analog ein positives für die Fortziehenden. Wie oben in Abschnitt 4.1 bereits angedeutet, könnten sich die Gründe für einen Umzug im Alter verschieben und die erwarteten Vorzeichen würden sich umkehren.

**Gesundheitswesen** Als Proxy für die Versorgung mit Gesundheitsdienstleistungen wurde die Zahl der aufgestellten Krankenhausbetten pro Kreis verwendet. Die Bettenzahl der in den Landeskrankenhausplan aufgenommenen Krankenhäuser wurde für 1999 aus einer Veröffentlichung des baden-württembergischen Sozialministeriums übernommen (Sozialministerium Baden-Württemberg 2003). Diese Vorgehensweise verursacht einen Messfehler, da einige Krankenhäuser es aufgrund der Unabhängigkeit von der staatlichen Bedarfsplanung vorziehen, nicht in den Plan aufgenommen zu werden. Eine komplette Statistik existiert beim Statistischen Landesamt Baden-Württemberg in Papierform, diese ist jedoch sehr detailliert und hätte einen überproportional hohen Bearbeitungsaufwand erfordert. Die Daten für 1999 wurden im gesamten Zeitraum 1995 bis 2002 verwendet, was aus zwei Gründen möglich erscheint. Erstens stellt die Bettenzahl eine kurzfristig fixe Größe dar, Neubauten oder Schließungen großer Kliniken dürften in dieser Zeit kaum vorgekommen sein. Zweitens ist die gesuchte Korrelation nicht zwischen Bettenzahl und Sterblichkeit, sondern zwischen medizinischer Versorgung und Sterblichkeit. Aufgrund des medizinisch-technischen Fortschritts ist die innerkreisliche Variation der Zahl der Betten

### 5.1. Datensatzbeschreibung und Hypothesenformulierung

über die Zeit kein sinnvoller Indikator für die Qualität der medizinischen Versorgung. Insofern erscheint es gerechtfertigt, von vornherein nur die räumliche Variation zu verwenden.

Um dem Fakt Rechnung zu tragen, dass Krankenhäuser in Stadtkreisen oft das Umland mitversorgen, wurden mehrere Spezifikationen der Bettenzahl berechnet. Für die Konstruktion der Variable *bett\_stadt* wurde in Landkreisen, welche direkt an einen Stadtkreis grenzen, die Summe der Bettenzahlen beider Kreise verwendet. Stadtkreise und Landkreise, die nicht an einen Stadtkreis grenzen, blieben unverändert. Diese Variable dürfte in etwa das Bettenangebot abbilden, welches den Bewohnern im Bedarfsfall zur Verfügung steht. Die zweite Spezifikation definiert Regionen als Stadtkreise und alle direkt angrenzenden Landkreise. Alle pro Region vorhandenen Betten wurden aufsummiert und in jedem darin enthaltenen Kreis verwendet, Landkreise ohne angrenzenden Stadtkreis blieben unverändert. Die resultierende Variable *bett\_reg* könnte ein Indikator für das Spezialisierungspotenzial und damit die Qualität der Gesundheitsversorgung sein. Alle Variablen zur Bettenzahl wurden sowohl roh als auch bezogen auf die Einwohnerzahl in den Regressionen verwendet. Die Hypothese ist eindeutig: Ein hohe Zahl aufgestellter Betten sollte die Sterblichkeit negativ beeinflussen. Dies könnte jedoch durch einen Selektionseffekt (s. Abschnitt 4.1) überlagert werden, so dass das Vorzeichen a priori unbestimmt bleibt.

**Flächennutzung/Siedlungsstruktur** Für die Jahre 1981, 1987, 1993, 1997 und 2001 stehen Daten zu den Anteilen der Gebäude- und Frei-, Verkehrs-, Wald- und Landwirtschaftsflächen an der Gesamtfläche der Kreise zur Verfügung. Die Kategorie „Sonstiges“ ist vernachlässigbar. Einen signifikanten direkten Effekt würde man nur bei der Verkehrsfläche erwarten, die einen Proxy für Umwelt- und Lärmbelastung durch den Straßenverkehr darstellt und einen positiven Effekt auf die Sterblichkeit haben sollte. Bei allen anderen Variablen dürften Selektionseffekte etwaige direkte Effekte dominieren, die Vorzeichen sind a priori unbestimmt.

**Umwelt** Daten zur Luftbelastung liegen in zwei Formen vor: Emissionen von Schwefeldioxid, Kohlenmonoxid und Staub für die Jahre 1990, 1995 und 2000 (Staubemissionen fehlen für 2000) und die Immissionen von Schwefeldioxid, Kohlenmonoxid und Ozon für die Jahre 1999-2002. Der sinnvollere Indikator ist die Immissionsbelastung der Luft, da Emissionen aufgrund von Luftströmungen und Klimaverhältnissen nicht zwangsläufig an der Quelle belastend wirken. Leider sind die Daten jedoch nur in einer relativ kurzen Zeitreihe vorhanden und das Netz von Umweltmessstationen deckt nicht alle Kreise ab. So fehlen der Landkreis Heilbronn, der Enzkreis, der Landkreis Emmendingen, der Zollernalbkreis und der Landkreis Sigmaringen in den Daten, hier liegt also die Struktur

## 5. Regressionsanalyse mit Paneldaten

eines *unbalanced panel* vor. Die Annahme eines von der Sterblichkeit unabhängigen Selektionsprozesses erscheint nicht besonders restriktiv, so dass diese Kreise ignoriert werden können, ohne dass die Schätzer inkonsistent werden. Es entsteht nur ein Problem aufgrund der geringeren Beobachtungszahl. Von den halbstündlichen Messungen der Luftbelastung wurden die Mittelwerte und das 98%-Quantil verwendet, um die durchschnittliche sowie die Spitzenbelastung abzubilden.

Aus den Emissionsdaten wurde durch Inter- und Extrapolation eine geschlossene Zeitreihe von 1990 bis 2002 gewonnen, im Falle von Staub nur bis 1995. Diese Vorgehensweise verstärkt einen möglichen Messfehler, der ohnehin schon erheblich sein dürfte, da beispielsweise die vom Straßenverkehr verursachten Belastungen nur geschätzt werden können. Die Hypothese für sämtliche Umweltbelastungen ist eindeutig, sie sollten positiv mit der Sterblichkeit korreliert sein. Etwaige Selektionsmechanismen, die auf sozioökonomischen Charakteristika beruhen (d.h. reiche und gut gebildete Personen ziehen in Kreise mit geringer Luftbelastung), könnten diesen Effekt noch verstärken.

**Sonstiges** In alle Spezifikationen wurden Dummyvariable für die Stadtkreise und sämtliche Jahre aufgenommen. Erstgenannte bilden mögliche Besonderheiten von Städten ab, die nicht von den anderen Regressoren erfasst werden. Ihr Vorzeichen ist a priori unklar. Die Jahresdummies bilden den landesweiten Durchschnitt der jährlichen Mortalitätsveränderung ab, da der Versuch einer Erklärung der durchschnittlichen Sterblichkeitsverbesserung mit den vorliegenden Daten nicht sinnvoll erscheint, weil hier komplizierte Lagstrukturen, qualitative Variable wie Lebensgewohnheiten und externe Einflüsse wie Kriegseffekte eine dominierende Rolle spielen dürften. Diese Probleme sind in der Querschnittsdimension weniger relevant, da sie sich im ganzen Land weitgehend homogen verändern dürften.

## 5.2. Ergebnisse

In nahezu allen Fällen lieferten die Random Effects-Schätzungen mit Stata gleichwertige Ergebnisse für die Gesamtbevölkerung und die deutsche Bevölkerung. Die meisten der im Anhang aufgeführten Tabellen beziehen sich auf die letztgenannte. Dies liegt zum einen an der zweifelhaften Qualität der Sterblichkeitsdaten für die ausländische Bevölkerung, zum anderen an der oftmals größeren Präzision der Schätzergebnisse. Falls die Resultate stark voneinander abwichen, wird im Text darauf hingewiesen. Vorab erscheinen einige Hinweise zur Interpretation der Regressionen angebracht.

- Die zum Teil sehr hohen Werte von  $R^2_{\text{within}}$  dürften vor allem durch die Jahresdummies zu Stande kommen, im Falle von  $R^2_{\text{between}}$  spielen die Momente der Altersverteilung eine große Rolle. Rückschlüsse auf den Erklärungsgehalt der relevanten Regressoren sollten daher nur mit besonderer Vorsicht gezogen werden.
- Die Linearitätsannahme ist bestenfalls als eine lokal erfüllte Approximation zu sehen, Interpretationen der Koeffizienten sollten sich daher auf Vorzeichen und Signifikanz beschränken.
- Falls Variable Verwendung fanden, die nicht nach Geschlecht differenziert vorlagen, so wurden zusätzlich entsprechend des Ansatzes von Rosenbaum und Rubin (1984) Modelle für die Gesamtbevölkerung berechnet, wobei die Frauenquote als Regressor aufgenommen wurde. Die Ergebnisse waren denen der stratifizierten Regressionen sehr ähnlich, so dass eine separate Auflistung der Resultate nicht gerechtfertigt erschien.

Der verbleibende Teil dieses Kapitels beschreibt die Regressionsergebnisse und orientiert sich dabei an den einzelnen Wirkungskomplexen. Eine Zusammenfassung der Ergebnisse findet sich in Tabelle 5.2. Hier wird gleich deutlich, dass sich die präzisesten Ergebnisse für die Einkommensgrößen, beschäftigte Hochschulabsolventen am Wohnort, die Quote der sozialversicherungspflichtig Beschäftigten am Wohnort und die Wanderungsbewegungen bei Frauen im Alter ergeben. Die folgenden Abschnitte befassen sich näher mit den einzelnen Einflussfaktoren.

Ausführliche Darstellungen ausgewählter Regressionen beinhaltet der Anhang in den Tabellen A.15 bis A.29. Die Spezifikationen orientieren sich dabei stark an der in Tabelle 5.1 abgebildeten Datenverfügbarkeit. Tabellen A.15 bis A.19 basieren auf einem Panel mit der Zeitdimension 1999 bis 2002, für welche die sinnvollsten Daten vorliegen. Allerdings ist die Zahl der Beobachtungen gering. Die folgenden Tabellen A.20 bis A.22 nutzen insbesondere die Verfügbarkeit der Einkommensteuerstatistik aus, Tabellen A.23 bis A.25 orientieren sich am Vorhandensein der Krankenhaus- und Sozialhilfestatistiken. Die Regressionen in den Tabellen A.26 und A.27 versuchen, einige Effekte anhand möglichst großer Zeitdimensionen herauszuarbeiten. Abschließend finden sich zwei einfache Spezifikationen mit den Immissionsvariablen in den Tabellen A.28 und A.29. Diese sind die einzigen Tabellen, in denen nicht die Struktur von *balanced panels* vorhanden ist.

**Sozioökonomische Variable** Für die Zeitreihen 1983 bis 1995 und 1999 bis 2002 liegen sowohl Bildungs- als auch Einkommensvariable vor, insofern bieten die Tabellen A.15 bis

## 5. Regressionsanalyse mit Paneldaten

Tabelle 5.2.: Zusammenfassung der Regressionsergebnisse: Vorzeichen und Signifikanz

	Alle Alter		Über 65	
	Männer	Frauen	Männer	Frauen
Einkommen (VDR)	(-) <sup>0</sup>	(-) <sup>***</sup>	(-) <sup>**</sup>	(-) <sup>***</sup>
Einkommen (EkSt)	(-) <sup>**</sup>	(-) <sup>0</sup>	(-) <sup>***</sup>	(-) <sup>0</sup>
Uni, Wohnort	(-) <sup>**</sup>	(-) <sup>***</sup>	(+) <sup>*</sup>	(+) <sup>*</sup>
Uni, Arbeitsort	(-) <sup>0-***</sup>	(-) <sup>0</sup>	(-) <sup>0</sup>	(-) <sup>0</sup>
Ausbildung, Wohnort	(+) <sup>0-***</sup>	(-) <sup>0-***</sup>	(+) <sup>0</sup>	(-) <sup>*</sup>
Ausbildung, Arbeitsort	(-) <sup>0-***</sup>	(-) <sup>0</sup>	(-) <sup>0</sup>	(+) <sup>0</sup> / (-) <sup>**</sup>
Beschäftigte <sup>1</sup> , Wohnort	(+) <sup>***</sup>	(+) <sup>**</sup>	(+) <sup>*</sup>	(+) <sup>**</sup>
Beschäftigte <sup>1</sup> , Arbeitsort	(+) <sup>0-***</sup> / (-) <sup>0</sup>	(-) <sup>0</sup>	(+) <sup>0</sup>	(-) <sup>0</sup>
Erwerbsminderungsrenten	(+) <sup>*</sup>	(+) <sup>0</sup>	(+) <sup>*</sup>	(+) <sup>**</sup>
Arbeitslosenunterstützung	(+) <sup>0</sup>	(-) <sup>0</sup>	(+) <sup>0</sup>	(+) <sup>0</sup>
Sozialhilfeempfänger	(+) <sup>**</sup>	(+) <sup>*</sup>	(+) <sup>0</sup>	(+) <sup>0</sup>
Zuwanderung	(-) <sup>0-***</sup>	(+) / (-)	(+) <sup>0</sup>	(+) <sup>***-***</sup>
Abwanderung	(+) <sup>0-*</sup>	(+) / (-)	(-) <sup>0</sup>	(-) <sup>***-***</sup>
Krankenhausbetten	(-) <sup>**</sup>	(-) <sup>*</sup>	(+) <sup>0</sup>	(-) <sup>0</sup>
Siedlungsstruktur	(+) / (-)	(+) / (-)	(+) / (-)	(+) / (-)
Umwelt	(+) / (-)	(+) / (-)	(+) / (-)	(+) / (-)

*Quelle:* Statistisches Landesamt Baden-Württemberg, VDR, Sozialministerium Baden-Württemberg, eigene Berechnungen.

<sup>1</sup>Sozialversicherungspflichtig Beschäftigte.

<sup>0</sup> insignifikant, \* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

A.22 die besten Datenstrukturen zur Überprüfung des Einflusses der sozioökonomischen Faktoren. Für die letzten Jahre des Untersuchungszeitraums können die Bildungsvariablen am Wohnort differenziert nach Geschlecht verwendet werden, welche die beste vorhandene Messgröße für das Bildungsniveau darstellen. Dasselbe gilt auch für das Einkommen im Falle der Alterssterblichkeit. Für jüngere Personen ist unklar, ob die Probleme der Einkommensteuerstatistik oder die Verwendung des Renteneinkommens als Indikator für das gesamte Kreiseinkommen größere Messfehler mit sich bringen.

Tabelle A.15 beinhaltet die Einflüsse der Bildungsvariablen und des Einkommens auf die Gesamtsterblichkeit deutscher Männer, wobei als weitere Kontrollvariable die Erwerbstätigenquote und Stadtdummys verwendet wurden. Das Einkommen hat einen insignifikanten sterblichkeitsmindernden Effekt. Dasselbe Vorzeichen ergibt sich auch für die Quote der Arbeitnehmer in der höchsten Bildungsklasse, der Koeffizient ist signifikant auf dem 5%-Niveau. Ein insignifikanter Effekt mit positivem Vorzeichen ergibt sich für die Quote der Arbeitnehmer mit abgeschlossener Berufsausbildung, letzteres entspricht nicht den Erwartungen. Die Erwerbstätigenquote hat einen signifikant sterblichkeitserhöhenden Einfluss. Demnach herrscht in der Gruppe der nicht abhängig Beschäftigten eine geringere Sterblichkeit vor: Beamte, Selbständige und Studenten scheinen stärker ins Gewicht zu fallen als Arbeitslose und Sozialhilfeempfänger. Hinzufügen oder Auslassen von einzelnen Regressoren änderte nichts Wesentliches am Ergebnis, gleiches gilt für die Betrachtung der Gesamtbevölkerung.

Die folgende Tabelle A.16 betrachtet nur die Sterblichkeit der Bevölkerung im arbeitsfähigen Alter. Die höchste Bildungskategorie und das Einkommen tauschen die Signifikanz, in Städten findet sich eine höhere Sterblichkeit als in den Landkreisen. In dieser Altersgruppe lassen sich signifikant positive Effekte des Anteils der Bezieher von Erwerbsminderungsrenten und derer von Arbeitslosenunterstützung nachweisen. Beides entspricht den Erwartungen. Tabelle A.17 enthält die entsprechenden Regressionsergebnisse für die Alterssterblichkeit. Auch hier lässt sich eine signifikant negative Korrelation von Einkommen und Mortalität feststellen. Entgegen den Vorhersagen und den bisherigen Resultaten ist das Vorzeichen beider Bildungsindikatoren positiv, zum Teil sogar leicht signifikant. Mögliche Erklärungen liegen in der ungerechtfertigten Übertragung der Bildungsindikatoren aus der arbeitenden Bevölkerung oder in einer ungenügenden Berücksichtigung der Altersverteilung (die Verwendung höherer Momente änderte nichts am Ergebnis). Beim Vergleich der drei Regressionen sticht vor allem die Tatsache ins Auge, dass das Einkommen in den letzten beiden Spezifikationen einen signifikanten Einfluss hat, jedoch nicht in der ersten. Diese stellt jedoch einen Oberfall im Vergleich zur Stratifizierung nach Alter dar. Falls das lineare Modell tatsächlich gelten sollte und die Kindersterblichkeit keinen signifi-



## 5. Regressionsanalyse mit Paneldaten

kanten Einfluss ausübt, so dürfte eine solche Konstellation nicht auftreten. Möglicherweise ist wiederum die Abbildung der Altersverteilung nicht ausreichend.

In den Tabellen A.18 und A.19 finden sich die entsprechenden Analysen für die Gesamt- und Alterssterblichkeit deutscher Frauen. Die erste Spezifikation liefert exakt die erwarteten Ergebnisse. Zunächst findet sich eine signifikant negative Korrelation von Einkommen sowie höherer Bildung und Sterblichkeit. Der Koeffizient des Anteils der Personen mit niedrigem Bildungsabschluss ist ebenfalls negativ, jedoch kleiner und insignifikant. Wie schon bei der männlichen Bevölkerung ist die Sterblichkeit der sozialversicherungspflichtig Beschäftigten höher als diejenige der restlichen Bevölkerung, zwischen Stadt und Land finden sich keine signifikanten Unterschiede. Im Fall der in Tabelle A.19 aufgeführten Einflüsse auf die Alterssterblichkeit ist als erheblicher Unterschied die Umkehrung der Vorzeichen der Bildungsvariablen zu nennen. Mögliche Ursachen sind dieselben wie bei den Männern. Der Anteil der Bezieher von Erwerbsminderungsrenten ist auch hier signifikant und positiv. Da diese nur für die 18 bis 65-jährigen erfasst werden, deutet dies darauf hin, dass sich die Verteilung von gesundheitsschädigenden Berufen über die Kreise im Zeitablauf nur wenig ändert. Eine Analyse der weiblichen Mortalität zwischen 18 und 65 Jahren hielt keine neuen Erkenntnisse bereit, auch hier ergibt sich ein positiv signifikanter Koeffizient der Arbeitslosenquote.

Die Tabellen A.20 bis A.22 beziehen sich auf Zeitreihen der Gesamtsterblichkeit von 1983 bis 1995 und beinhalten als wesentlichen Bestandteil die Daten der Einkommensteuerstatistik. Die Spezifikation des Mittelwerts mit der oberen Grenze des Einkommens bei 125.000 DM erwies sich als die sinnvollste Alternative, insbesondere gegenüber derjenigen mit 200.000 DM. Letztere lieferte in einer sehr viel geringeren Zahl der Fälle signifikante Koeffizienten. Dies spricht für einen abnehmenden Steigungsparameter des Zusammenhangs von Einkommen und Sterblichkeit, was auch intuitiv Sinn macht. Das so bestimmte mittlere Kreiseinkommen übt bei Männern aller Altersklassen einen signifikant sterblichkeitsverringenden Einfluss aus. Die Bildungsindikatoren sind sowohl für die Gesamt- als auch für die Alterssterblichkeit negativ und insignifikant. Letzteres könnte daran liegen, dass die Bildungsvariablen nur am Arbeitsort vorliegen. Dies wirkt sich möglicherweise auch auf die Erwerbsquote aus, welche in beiden Fällen ein negatives Vorzeichen hat, ohne jedoch signifikant zu sein. Vergleiche mit den Resultaten für am Wohnort gemessene Variable sollten daher nicht gezogen werden.

Die in Tabelle A.22 aufgeführten Ergebnisse gelten exemplarisch für sämtliche Regressionen auf die weibliche Sterblichkeit, welche auf dieser Zeitreihe beruhen: Während die Koeffizienten die erwarteten Vorzeichen besitzen, bleiben sie in allen Fällen insignifikant. Der wesentliche Grund dürfte in der fehlenden Geschlechtsdifferenzierung der Einkommens-

und Bildungsindikatoren liegen. Beide beruhen weitgehend auf der erwerbstätigen Bevölkerung. In dieser Gruppe sind Männer jedoch überrepräsentiert, so dass auch die Indikatoren stärker männliche als weibliche Charakteristika abbilden. Die Insignifikanz sollte entsprechend kein Anlass zu vergleichenden Interpretationen sein. Gleiches gilt auch dafür, dass die Varianz des Einkommens wechselnde Vorzeichen aufwies und in keiner Spezifikation signifikant wurde. Die Daten der Einkommensteuerstatistik sind vermutlich am unteren Rand zu lückenhaft, als dass ein hinreichend gutes Streuungsmaß vorläge. Gegen die Ungleichheitshypothese von Wilkinson (1996) spricht jedoch, dass auch die Varianzspezifikationen mit hohen Obergrenzen des Einkommens keine signifikant positiven Effekte zeigten.

Anhand der Tabellen A.23 bis A.25 kann der Einfluss von Sozialhilfeempfängerquoten überprüft werden. Wie erwartet ergibt sich ein positives Vorzeichen, welches nur im Falle der Gesamtsterblichkeit signifikant ist. Der Anteil der Empfänger von Arbeitslosenunterstützung weist nur im Falle der männlichen Sterblichkeit zwischen 18 und 65 Jahren (nicht aufgeführt) einen signifikant positiven Koeffizienten auf. Hinsichtlich der sozioökonomischen Variablen ruft der signifikant positive Einfluss der mittleren Bildungskategorie in Tabelle A.23 Erstaunen hervor. Die Theorie bietet hierfür wenig Erklärungen, vermutlich liegt es an dem Messfehler, der durch die Erfassung der Indikatoren am Arbeitsort auftritt. Das negative Vorzeichen der Arbeitslosenquote in Tabelle A.24 könnte erneut durch die fehlende Differenzierung des Indikators nach dem Geschlecht hervorgerufen werden. Die Bildungsvariablen in den langen Zeitreihen der Tabellen A.26 und A.27 haben wiederum die erwarteten Vorzeichen und sind weitgehend signifikant.

Zusammenfassend lässt sich festhalten, dass sowohl das Einkommen als auch höhere Bildung einen sterblichkeitsmindernden Einfluss ausüben. Die Ergebnisse sind nicht immer einheitlich, was aber zum Großteil an der Datenqualität liegen dürfte. Für die mittlere Bildungskategorie lassen sich keine Effekte feststellen. Die Erwerbstätigenquote am Wohnort wies immer ein positives Vorzeichen auf, was auf eine höhere Sterblichkeit der sozialversicherungspflichtig Beschäftigten relativ zu den anderen Personengruppen hindeutet. Wurde sie am Arbeitsort erfasst, so finden sich wechselnde Vorzeichen und Signifikanzniveaus. Die Quote der Empfänger von Arbeitslosenunterstützung ist nur im Fall der Bevölkerung im arbeitsfähigen Alter signifikant. Der Anteil der Sozialhilfeempfänger ist in allen Altersgruppen positiv mit der Sterblichkeit korreliert, jedoch nicht in allen Fällen signifikant. Ein solcher Zusammenhang findet sich auch für die Quote der Bezieher von Erwerbsminderungsrenten.

**Wanderung** Die interessantesten Effekte von Wanderungsbewegungen finden sich in den Tabellen A.26 und A.27. Für die männliche Sterblichkeit findet sich ein sterblichkeits-

## 5. Regressionsanalyse mit Paneldaten

mindernder Effekt von Zuwanderungen und ein gegenteiliger von Abwanderungen. Dies entspricht den Erwartungen insofern, als dass mobile Personen vergleichsweise gesund sein dürften. Die Variablen scheinen relativ stark mit der Bildung korreliert zu sein, beide steigen deutlich sowohl hinsichtlich der Größe als auch der Signifikanz, wenn das Bildungsniveau keine Berücksichtigung in der Regression findet. Wird die Bildung am Wohnort konstant gehalten, so sind die Wanderungsbewegungen insignifikant, allerdings liegt in diesem Fall nur eine sehr kleine Zeitdimension zugrunde.

Diese Selektionseffekte scheinen bei Frauen weniger ausgeprägt zu sein, im Fall der Gesamtsterblichkeit und jener zwischen 18 und 65 Jahren bleiben alle Wanderungsbewegungen bei wechselnden Vorzeichen insignifikant. Dabei macht es keinen Unterschied, ob die Bildung in der Regression berücksichtigt wird. Um so stärker schlagen die Effekte in umgekehrter Richtung für die Alterssterblichkeit zu Buche (Tabelle A.27): Die Abwanderungen haben einen negativen, die Zuwanderungen einen positiven Effekt auf die Mortalität, beide sind hochsignifikant. Die Umkehrung der Vorzeichen lässt sich meist auch bei Männern beobachten, die Effekte bleiben jedoch insignifikant. Diese Befunde decken sich mit der Intuition, dass Umzüge in dieser Altersklasse vor allem aufgrund von einsetzender Pflegebedürftigkeit getätigt werden. Da Ehemänner aufgrund der geringeren Lebenserwartung häufiger im eigenen Haushalt gepflegt werden können, ist der Effekt für Frauen deutlich stärker ausgeprägt.

**Gesundheitswesen** Den Effekt des Gesundheitswesens auf die Sterblichkeit untersuchen die Tabellen A.23 bis A.25. Von allen sechs getesteten Spezifikationen wies nur die Zahl der Betten pro „Region“ eine signifikant negative partielle Korrelation auf, und zwar nur für die Gesamtsterblichkeit. Dies könnte als Indiz dafür gewertet werden, dass eine hohe regionale Konzentration von Betten zu einer deutlichen Verbesserung der medizinischen Versorgung führt, möglicherweise aufgrund des erhöhten Spezialisierungspotenzials oder erleichtertem Wissenstransfer. Die Ergebnisse stehen jedoch auf wenig soliden Fundamenten, für solch weitgehende Interpretationen wären weitere Kontrollvariable zum Gesundheitswesen notwendig.

Die anderen Spezifikationen lieferten keine präzisen Resultate. Die ursprüngliche rohe Bettenzahl wies ein positives Vorzeichen auf, die Variable *bett\_stadt* ein negatives. Beide blieben in allen Fällen insignifikant. Die auf die Einwohnerzahl bezogenen Bettenzahlen waren bei wechselnden Vorzeichen in allen Varianten meist insignifikant. Nur bei der Alterssterblichkeit traten einige signifikant positive Koeffizienten auf. Diese deuten auf einen Selektionseffekt dahingehend hin, dass gesundheitlich beeinträchtigte Personen die Nähe

von Kliniken suchen. Die umgekehrte Richtung wäre ebenfalls denkbar: Eine hohe Bettenzahl könnte die Reaktion auf eine vergleichsweise ungesunde Bevölkerung sein.

**Flächennutzung** Informationen über den Einfluss der Siedlungsstruktur finden sich in den Tabellen A.23 bis A.27. Die Kategorien Landwirtschaftsfläche und Sonstiges sind ausgelassen. Der Anteil der Verkehrsfläche hat in aller Regel ein positives Vorzeichen, ist jedoch nur selten signifikant. Aus der Reihe fällt nur das negative Vorzeichen in Tabelle A.25, was aufgrund der hohen Standardabweichung jedoch nicht überbewertet werden sollte. Auf den ersten Blick verwunderlich ist auch der positive Effekt der Waldfläche, welcher in den längeren Zeitreihen auch signifikant ist. Hier dürfte jedoch ein Selektionseffekt vorliegen und mangelnde Abbildung der sozioökonomischen Charakteristika eher eine Rolle spielen, als ein direkter Effekt von Wäldern auf die Sterblichkeit. Der Anteil der Gebäude- und Freifläche liefert keine einheitlichen Ergebnisse.

**Umwelt** Exemplarisch für den Einfluss von Luftbelastungen sind in den Tabellen A.28 und A.29 die Ergebnisse der Auswertungen zu den Immissionsdaten aufgeführt. Es machte keinen Unterschied, ob die sozioökonomischen Charakteristika in die Regression aufgenommen wurden oder nicht. A priori würde man einen negativen Selektionseffekt erwarten, so dass negative Einflüsse der Luftbelastung ohne Berücksichtigung der sozioökonomischen Variablen höher erscheinen dürften. Obwohl dies in den aufgeführten Tabellen der Fall war, findet sich keinerlei Einfluss der Umweltbelastungen. Das 98%-Quantil der Kohlenmonoxidbelastung scheint sogar eine negative Korrelation mit der Mortalität aufzuweisen. Sämtliche anderen Spezifikationen sowohl der Immissions- als auch der Emissionsdaten wiesen ebenfalls keine signifikanten Effekte auf, unabhängig von den verwendeten Kovariaten. Auch diese Ergebnisse sollten nicht überbewertet werden, da alle Variable die tatsächliche Exposition von Individuen mit Luftbelastungen nur unzureichend abbilden. Insbesondere im Fall von Ozon reicht die Verwendung des 98%-Quantils vermutlich nicht aus, um die medizinisch relevante Spitzenbelastung abzubilden. Zudem ist die Zeitreihe der Immissionsstatistik sehr kurz, die Werte der Emissionen sind mit mehreren Messfehlern behaftet.

**Sonstiges** Der Stadtdummy ist für die Gesamtsterblichkeit in nahezu allen Fällen positiv, unabhängig von den gewählten Kovariaten und Zeitpunkten. Interessant ist sein Verhalten im Falle der Alterssterblichkeit, da in Abschnitt 2.3 bereits eine relativ bessere Position der Stadtkreise für die fiktive fernere Lebenserwartung im Alter 65 festgestellt wurde. Tatsächlich weist der Stadtdummy ein negatives Vorzeichen auf, wenn nicht auf sozioökonomische Variable konditioniert wird (die entsprechenden Regressionen sind nicht aufgeführt).

## 5. Regressionsanalyse mit Paneldaten

Falls jedoch geschlechtsspezifische Bildungsindikatoren in der Regression enthalten sind, so wechselt das Vorzeichen und der Koeffizient ist nur selten signifikant. Entsprechend scheinen keine systematischen Unterschiede zwischen Stadt und Land vorhanden zu sein, die geringere Alterssterblichkeit in den Stadtkreisen erklärt sich offensichtlich durch sozioökonomische Charakteristika.

Bei den Regressionen für die Gesamtbevölkerung erwies es sich als entscheidend, den Ausländeranteil als Regressor aufzunehmen. Seine Nichtberücksichtigung hatte insbesondere starke Veränderungen bei Vorzeichen und Signifikanz der Bildungsvariablen zur Folge. Wie erwartet hatte die Ausländerquote in allen Fällen ein negatives Vorzeichen, die Signifikanz schwankte je nach Spezifikation.

Der Erklärungsgehalt der Regressionen zur Gesamtsterblichkeit ist ausnahmslos sehr hoch, das für diese Analyse besonders interessante  $R^2_{\text{between}}$  liegt bei rund 0,9. Wie oben bereits angedeutet, liegt dies jedoch wesentlich an den Momenten der Altersverteilung. Durch die Verwendung der sozioökonomischen Charakteristika in Tabelle A.15 anstelle der Immissionen in Tabelle A.28 steigt es beispielsweise nur um 0,04. Die Varianz der Störterme erklärt sich zu jeweils rund der Hälfte aus zeitkonstanten und zeitvariablen Effekten. Bei der Alterssterblichkeit liegt das  $R^2_{\text{between}}$  zwischen 0,2 und 0,5, die zeitkonstanten un beobachteten Effekte haben einen geringeren Einfluss auf die Varianz der Störterme.

## 6. Zusammenfassung und Ausblick

Diese Arbeit versuchte, Determinanten der Sterblichkeit auf aggregierter Ebene herauszuarbeiten. Ausgangspunkt waren die im zweiten Kapitel dargestellten regionalen Mortalitätsunterschiede in Baden-Württemberg auf Kreisebene. Der folgende Teil der Arbeit zeigte, dass Zufallsschwankungen allein als Erklärung nicht ausreichen. Ein signifikanter Einfluss der in der amtlichen Statistik zu erwartenden Fehler konnte ebenfalls ausgeschlossen werden. Interessant wäre eine weitergehende Analyse, inwieweit Messfehler für die deutlich höhere Lebenserwartung der Ausländer verantwortlich sein könnten. Zwar ist eine Übersterblichkeit der einheimischen Bevölkerung aufgrund von Selektionseffekten zu erwarten, die vorgefundenen Unterschiede sind jedoch außerhalb der vorstellbaren Bereiche. Aufgrund der oftmals geringen Ausländerzahlen in den Kreisen müsste eine solche Arbeit auf einer höheren Aggregationsebene erfolgen.

Nach dem Ausschluss dieser trivialen Ursachen wurde ein knapper Überblick der theoretischen und empirischen Literatur zu Mortalitätsunterschieden gegeben. Dabei lag ein besonderes Augenmerk auf sozioökonomischen Faktoren. Anschließend wurde auf die Analyse aggregierter Daten eingegangen und ein einfaches ökologisches Regressionsmodell vorgestellt, welches auf die Datenlage zugeschnitten ist. Aufgrund der notwendigen Linearitätsannahme sollten sich die Interpretationen auf qualitative Effekte beschränken.

In Kapitel 5, welches den Hauptteil der Arbeit bildet, wurde dieses Modell in verschiedenen Spezifikationen angewendet. Schlagwortartig lauten die Ergebnisse wie folgt:

- Der Zusammenhang zwischen Einkommen und Mortalität ist signifikant negativ.
- Gleiches gilt für höhere Bildung, zumindestens im Fall der Gesamtsterblichkeit. Die Ergebnisse für die Alterssterblichkeit sind uneinheitlich, was jedoch seine Ursache in Messfehlern haben könnte.
- Die Sterblichkeit von sozialversicherungspflichtig Beschäftigten ist höher als die der restlichen Personengruppen.

## 6. Zusammenfassung und Ausblick

- Ein hoher Arbeitslosenanteil ist positiv mit der Mortalität der Bevölkerung im arbeitsfähigen Alter korreliert, für die Gesamtbevölkerung oder ältere Menschen lassen sich keine Effekte nachweisen.
- Die Quoten der Bezieher von Erwerbsminderungsrenten und der Sozialhilfeempfänger weisen eine positive Korrelation mit der Sterblichkeit auf, die Signifikanzniveaus wechseln jedoch.
- Abwanderungen weisen eine positive, Zuwanderungen eine negative, Korrelation mit der Gesamtsterblichkeit auf. Sie schwächt sich jedoch deutlich ab, wenn für sozioökonomische Variablen kontrolliert wird. Ein umgekehrter und für Frauen hochsignifikanter Effekt ergibt sich für die Alterssterblichkeit, was an veränderten Gründen für Umzüge liegen dürfte.
- Bezüglich der Einrichtungen des Gesundheitswesens lassen sich keine einheitlichen Aussagen treffen.
- Es konnte kein Einfluss von Umweltbelastungen nachgewiesen werden. Angesichts der dünnen Datenlage und ungesicherter Expositionsraten sollte dieses Ergebnis jedoch nicht überbewertet werden.

Diese Ergebnisse stehen nicht im Widerspruch zu vorangegangenen Studien in Baden-Württemberg (Cischinsky 2000) und anderen Bundesländern (Wittwer-Backofen (1999), Henke und Müller (2002)), deren Resultate konnten jedoch aufgrund der besseren Datenlage hinsichtlich der sozioökonomischen Indikatoren zumindestens in dieser Beziehung verfeinert werden. Insbesondere konnte kein signifikanter Einfluss von politikrelevanten Faktoren wie Umweltbelastungen oder der Gesundheitsversorgung festgestellt werden. Dies bedeutet natürlich nicht, dass diese Variablen per se keinen Einfluss auf die Lebenserwartung haben. Aber offenbar sind in Baden-Württemberg Luftverschmutzung oder Gesundheitsversorgung auf Kreisebene so einheitlich, dass sie keine Auswirkungen auf die Gesamtsterblichkeit haben. Die Unterschiede scheinen sich überwiegend aus individuellen Verhaltensweisen und Charakteristika zu ergeben, auf die politische Maßnahmen nur wenig Einfluss nehmen können. Dies gilt auch für die Bildung: zwar kann die Politik Einfluss auf das Bildungsniveau einer Gesellschaft nehmen, allerdings hängt nur wenig von ihr ab, welchen Wohnort die Menschen im weiteren Verlaufe ihres Lebens wählen.

# A. Tabellen

Tabelle A.1.: Vergleich der Bevölkerung Baden-Württembergs

<b>Jahr</b>	<b>Deutsche Männer</b>	<b>Ausl. Männer</b>	<b>Ausländer- anteil</b>	<b>Deutsche Frauen</b>	<b>Ausl. Frauen</b>	<b>Ausländer- anteil</b>
1980	3940574	523278	11,72	4392047	403045	8,41
1981	3949649	527986	11,79	4395875	414408	8,62
1982	3956441	507486	11,37	4397323	409355	8,52
1983	3955776	491903	11,06	4393697	401922	8,38
1984	3975047	470848	10,59	4405602	389586	8,12
1985	3989102	476068	10,66	4412333	393867	8,19
1986	4008210	490382	10,90	4422049	406139	8,41
1987	4029485	482961	10,70	4429155	388887	8,07
1988	4058595	509555	11,15	4448988	415571	8,54
1989	4131960	538208	11,52	4505968	442560	8,94
1990	4209564	574088	12,00	4565643	472732	9,38
1991	4250233	631342	12,93	4597066	523199	10,22
1992	4274915	691682	13,93	4616455	565656	10,92
1993	4292544	715399	14,29	4629226	596857	11,42
1994	4307109	715476	14,25	4637256	612228	11,66
1995	4324198	722649	14,32	4647039	625481	11,86
1996	4344410	732310	14,42	4659688	638097	12,04
1997	4386686	701947	13,79	4693460	614517	11,58
1998	4412176	693357	13,58	4710036	610471	11,47
1999	4440873	691492	13,47	4729884	613683	11,48
2000	4481180	679701	13,17	4759093	611129	11,38
2001	4519243	679872	13,08	4786789	617039	11,42
1987-1986	21275	-7421		7106	-17252	
%	+0,53	-1,51		+0,16	-4,25	

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.



A. Tabellen

Tabelle A.2.: Vergleich der Bevölkerung Baden-Württembergs über 65 Jahren

<b>Jahr</b>	<b>Deutsche Männer</b>	<b>Ausl. Männer</b>	<b>Ausländer- anteil</b>	<b>Deutsche Frauen</b>	<b>Ausl. Frauen</b>	<b>Ausländer- anteil</b>
1980	474678	7944	1,65	832459	9928	1,18
1981	463475	8375	1,77	825072	10397	1,24
1982	450826	8584	1,87	815111	10566	1,28
1983	437934	8883	1,99	805201	10919	1,34
1984	436967	9342	2,09	813214	11489	1,39
1985	442919	10174	2,25	833814	12365	1,46
1986	449337	10969	2,38	852924	13235	1,53
1987	455421	8087	1,74	870455	8415	0,96
1988	459509	9154	1,95	883365	9678	1,08
1989	463358	10377	2,19	894824	11048	1,22
1990	472299	11617	2,40	907490	12756	1,39
1991	482194	13261	2,68	916574	15455	1,66
1992	494227	14826	2,91	924193	17120	1,82
1993	509927	16817	3,19	931889	19316	2,03
1994	525745	19051	3,50	938962	21696	2,26
1995	541929	21359	3,79	946133	23829	2,46
1996	554737	23956	4,14	949369	25666	2,63
1997	568214	25010	4,22	953554	25218	2,58
1998	579687	27891	4,59	955348	26961	2,74
1999	598466	31719	5,03	965072	29643	2,98
2000	622377	34850	5,30	979662	32064	3,17
2001	647840	38665	5,63	997232	34642	3,36
1987-1986	6084	-2882		17531	-4820	
%	+1,35	-26,27		+2,06	-36,42	

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.3.: Fiktive Lebenserwartung bei Geburt, Männer

Kreis	1981	2001	jährliche Änderung (%)	Mittelwert
Stadtkreis Stuttgart	72,45	77,71	0,35	74,58
Landkreis Böblingen	72,46	78,30	0,38	75,34
Landkreis Esslingen	72,46	77,68	0,34	75,06
Landkreis Göppingen	71,79	77,42	0,37	74,31
Landkreis Ludwigsburg	71,87	78,28	0,42	74,79
Rems-Murr-Kreis	72,09	77,27	0,34	74,68
Stadtkreis Heilbronn	71,85	77,59	0,38	74,05
Landkreis Heilbronn	70,80	76,22	0,37	74,04
Hohenlohekreis	73,24	77,14	0,25	74,13
Landkreis Schwäbisch Hall	71,93	76,31	0,28	73,65
Main-Tauber-Kreis	68,94	76,04	0,47	74,05
Landkreis Heidenheim	70,99	77,16	0,41	73,97
Ostalbkreis	70,55	77,08	0,43	73,67
Stadtkreis Baden-Baden	71,86	76,35	0,28	73,82
Stadtkreis Karlsruhe	71,15	77,26	0,41	73,50
Landkreis Karlsruhe	70,69	76,98	0,42	73,52
Landkreis Rastatt	69,47	76,10	0,45	73,39
Stadtkreis Heidelberg	68,93	76,65	0,51	74,28
Stadtkreis Mannheim	69,45	75,16	0,39	72,51
Neckar-Odenwald-Kreis	71,43	75,08	0,24	73,02
Rhein-Neckar-Kreis	70,90	77,36	0,43	73,97
Stadtkreis Pforzheim	70,73	77,98	0,47	73,79
Landkreis Calw	71,19	77,02	0,39	74,10
Enzkreis	72,39	76,69	0,28	74,53
Landkreis Freudenstadt	70,97	75,80	0,32	73,48
Stadtkreis Freiburg	71,31	77,22	0,39	74,43
Landkreis Breisgau-Hochschwarzwald	71,86	77,66	0,38	75,03
Landkreis Emmendingen	70,85	77,64	0,45	73,98
Ortenaukreis	71,74	76,49	0,32	73,79
Landkreis Rottweil	71,04	76,36	0,35	73,98
Schwarzwald-Baar-Kreis	70,78	75,92	0,34	74,33
Landkreis Tuttlingen	70,48	76,66	0,41	74,04
Landkreis Konstanz	72,67	76,90	0,28	74,73
Landkreis Lörrach	70,75	77,79	0,47	74,35
Landkreis Waldshut	73,57	77,04	0,22	74,68
Landkreis Reutlingen	71,91	77,69	0,38	74,57
Landkreis Tübingen	73,75	78,64	0,32	75,63
Zollernalbkreis	69,27	77,13	0,53	73,97
Stadtkreis Ulm	70,17	77,95	0,50	74,72
Alb-Donau-Kreis	70,67	77,07	0,43	74,22
Landkreis Biberach	71,22	76,65	0,36	74,00
Bodenseekreis	72,16	78,20	0,40	75,27
Landkreis Ravensburg	70,50	77,58	0,47	74,22
Landkreis Sigmaringen	73,06	75,98	0,17	74,28
Mittelwert	71,33	77,03	0,37	74,19
Minimum	68,93	75,08	0,41	72,51
Maximum	73,75	78,64	0,31	75,63
Korrelation mit Verbesserung	-0,77	0,33	-	-0,06
Korrelation mit Mittelwert	0,55	0,73	-	-

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.4.: Fiktive Lebenserwartung bei Geburt, Frauen

Kreis	1981	2001	jährliche Änderung (%)	Mittelwert
Stadtkreis Stuttgart	78,78	83,09	0,26	80,95
Landkreis Böblingen	78,79	83,02	0,26	80,90
Landkreis Esslingen	77,91	82,90	0,31	80,87
Landkreis Göppingen	77,22	81,77	0,28	80,28
Landkreis Ludwigsburg	77,96	83,61	0,35	80,88
Rems-Murr-Kreis	77,53	83,10	0,34	80,61
Stadtkreis Heilbronn	77,45	80,77	0,20	80,38
Landkreis Heilbronn	76,42	82,62	0,39	79,82
Hohenlohekreis	78,26	83,69	0,33	80,36
Landkreis Schwäbisch Hall	77,40	82,58	0,32	79,91
Main-Tauber-Kreis	76,94	82,43	0,34	80,36
Landkreis Heidenheim	77,83	83,23	0,33	80,26
Ostalbkreis	78,17	82,78	0,28	79,85
Stadtkreis Baden-Baden	76,83	81,44	0,24	79,88
Stadtkreis Karlsruhe	76,54	82,67	0,38	79,66
Landkreis Karlsruhe	77,40	82,37	0,31	79,81
Landkreis Rastatt	77,32	81,98	0,29	79,95
Stadtkreis Heidelberg	76,78	82,07	0,33	80,78
Stadtkreis Mannheim	76,44	81,36	0,31	79,17
Neckar-Odenwald-Kreis	76,11	81,52	0,33	79,23
Rhein-Neckar-Kreis	77,29	83,24	0,37	80,06
Stadtkreis Pforzheim	77,89	81,34	0,20	80,41
Landkreis Calw	78,08	82,30	0,26	80,13
Enzkreis	79,06	82,54	0,21	80,75
Landkreis Freudenstadt	77,37	82,06	0,29	79,97
Stadtkreis Freiburg	77,47	82,76	0,33	80,88
Landkreis Breisgau-Hochschwarzwald	79,40	82,35	0,18	81,15
Landkreis Emmendingen	77,15	82,46	0,33	80,52
Ortenaukreis	77,33	82,71	0,33	80,39
Landkreis Rottweil	76,97	83,03	0,37	80,39
Schwarzwald-Baar-Kreis	78,13	82,19	0,25	80,47
Landkreis Tuttlingen	76,64	82,20	0,34	80,20
Landkreis Konstanz	78,55	82,82	0,26	80,94
Landkreis Lörrach	79,08	83,21	0,25	80,83
Landkreis Waldshut	78,95	83,36	0,26	80,94
Landkreis Reutlingen	78,37	83,70	0,33	80,94
Landkreis Tübingen	78,05	83,80	0,35	81,30
Zollernalbkreis	79,09	83,45	0,26	80,50
Stadtkreis Ulm	77,78	83,59	0,35	81,12
Alb-Donau-Kreis	78,15	82,87	0,29	80,53
Landkreis Biberach	78,16	82,99	0,29	80,22
Bodenseekreis	78,25	83,70	0,33	81,35
Landkreis Ravensburg	78,07	83,66	0,34	80,76
Landkreis Sigmaringen	77,95	81,68	0,22	80,45
Mittelwert	77,76	82,66	0,29	80,43
Minimum	76,11	80,77	0,28	79,17
Maximum	79,40	83,80	0,26	81,35
Korrelation mit Verbesserung	-0,56	0,46	-	-0,15
Korrelation mit Mittelwert	0,67	0,56	-	-

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.5.: Fiktive fernere Lebenserwartung im Alter 65, Männer

Kreis	1981	2001	jährliche Änderung (%)	Mittelwert
Stadtkreis Stuttgart	14,42	17,86	1,05	15,82
Landkreis Böblingen	13,88	17,51	1,10	15,62
Landkreis Esslingen	14,38	17,24	0,88	15,75
Landkreis Göppingen	14,54	17,18	0,80	15,48
Landkreis Ludwigsburg	13,77	17,51	1,18	15,50
Rems-Murr-Kreis	14,04	17,19	0,95	15,46
Stadtkreis Heilbronn	14,06	17,92	1,08	15,43
Landkreis Heilbronn	13,56	16,39	0,88	14,99
Hohenlohekreis	14,04	17,20	0,96	15,52
Landkreis Schwäbisch Hall	14,16	16,76	0,75	15,27
Main-Tauber-Kreis	14,12	16,90	0,84	15,33
Landkreis Heidenheim	13,59	17,31	1,14	15,31
Ostalbkreis	13,91	16,89	0,93	15,14
Stadtkreis Baden-Baden	14,24	16,86	0,78	15,45
Stadtkreis Karlsruhe	13,85	16,87	0,95	15,06
Landkreis Karlsruhe	12,89	16,81	1,27	14,71
Landkreis Rastatt	14,32	16,76	0,73	15,04
Stadtkreis Heidelberg	13,84	16,99	0,91	15,52
Stadtkreis Mannheim	13,06	16,06	1,00	14,69
Neckar-Odenwald-Kreis	13,71	16,02	0,70	14,63
Rhein-Neckar-Kreis	13,50	16,85	1,08	14,93
Stadtkreis Pforzheim	14,13	17,75	1,08	15,34
Landkreis Calw	13,65	17,13	1,10	15,37
Enzkreis	13,94	16,69	0,85	15,37
Landkreis Freudenstadt	14,35	16,14	0,47	14,98
Stadtkreis Freiburg	14,40	17,66	0,97	15,87
Landkreis Breisgau-Hochschwarzwald	14,68	17,54	0,84	16,07
Landkreis Emmendingen	14,39	17,69	0,96	15,43
Ortenaukreis	13,50	16,79	1,06	15,23
Landkreis Rottweil	13,95	17,11	0,95	15,26
Schwarzwald-Baar-Kreis	14,34	16,65	0,72	15,39
Landkreis Tuttlingen	13,30	17,26	1,24	15,39
Landkreis Konstanz	14,34	17,19	0,88	15,71
Landkreis Lörrach	14,09	17,28	0,97	15,34
Landkreis Waldshut	15,04	17,91	0,79	15,74
Landkreis Reutlingen	14,66	17,10	0,73	15,62
Landkreis Tübingen	14,80	17,85	0,88	15,77
Zollernalbkreis	13,92	17,17	0,99	15,27
Stadtkreis Ulm	14,56	17,79	0,87	15,91
Alb-Donau-Kreis	14,19	17,09	0,89	15,13
Landkreis Biberach	13,99	16,91	0,92	15,27
Bodenseekreis	14,71	17,51	0,84	15,97
Landkreis Ravensburg	13,87	17,26	1,02	15,64
Landkreis Sigmaringen	13,91	16,16	0,53	15,43
Mittelwert	14,06	17,11	0,94	15,39
Minimum	12,89	16,02	1,04	14,63
Maximum	15,04	17,92	0,84	16,07
Korrelation mit Verbesserung	-0,54	0,43	-	-0,04
Korrelation mit Mittelwert	0,70	0,73	-	-

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.6.: Fiktive fernere Lebenserwartung im Alter 65, Frauen

Kreis	1981	2001	jährliche Änderung (%)	Mittelwert
Stadtkreis Stuttgart	18,54	21,10	0,64	19,83
Landkreis Böblingen	17,96	20,68	0,68	19,12
Landkreis Esslingen	17,48	20,65	0,81	19,19
Landkreis Göppingen	17,55	20,28	0,70	18,79
Landkreis Ludwigsburg	17,15	21,00	0,99	19,15
Rems-Murr-Kreis	17,44	20,82	0,87	19,00
Stadtkreis Heilbronn	18,06	20,10	0,50	19,39
Landkreis Heilbronn	16,45	20,09	0,98	18,18
Hohenlohekreis	17,57	20,89	0,84	18,86
Landkreis Schwäbisch Hall	17,58	20,23	0,66	18,52
Main-Tauber-Kreis	17,10	20,31	0,83	18,83
Landkreis Heidenheim	16,76	20,59	0,98	18,73
Ostalbkreis	17,56	20,67	0,80	18,63
Stadtkreis Baden-Baden	17,17	20,54	0,84	19,01
Stadtkreis Karlsruhe	16,39	20,27	1,03	18,63
Landkreis Karlsruhe	16,28	20,15	1,04	18,18
Landkreis Rastatt	16,98	20,30	0,87	18,60
Stadtkreis Heidelberg	17,32	20,75	0,87	19,43
Stadtkreis Mannheim	17,05	19,81	0,73	18,56
Neckar-Odenwald-Kreis	16,14	19,29	0,84	17,81
Rhein-Neckar-Kreis	16,75	20,27	0,93	18,49
Stadtkreis Pforzheim	17,65	20,54	0,71	19,09
Landkreis Calw	17,39	20,11	0,69	18,60
Enzkreis	17,90	19,80	0,48	18,84
Landkreis Freudenstadt	17,02	20,18	0,83	18,25
Stadtkreis Freiburg	17,75	21,15	0,85	19,68
Landkreis Breisgau-Hochschwarzwald	17,94	20,41	0,60	19,44
Landkreis Emmendingen	16,50	20,55	1,07	18,87
Ortenaukreis	17,08	20,40	0,87	18,88
Landkreis Rottweil	16,80	20,25	0,89	18,99
Schwarzwald-Baar-Kreis	17,34	20,57	0,82	18,91
Landkreis Tuttlingen	16,66	20,67	1,05	18,94
Landkreis Konstanz	17,78	20,30	0,64	19,41
Landkreis Lörrach	17,91	20,85	0,71	19,13
Landkreis Waldshut	17,56	20,71	0,79	19,26
Landkreis Reutlingen	17,67	21,03	0,85	19,32
Landkreis Tübingen	17,52	20,73	0,82	19,34
Zollernalbkreis	17,20	20,52	0,86	18,94
Stadtkreis Ulm	17,27	21,12	0,96	19,60
Alb-Donau-Kreis	16,89	20,19	0,85	18,52
Landkreis Biberach	17,58	20,43	0,73	18,66
Bodenseekreis	18,09	20,95	0,71	19,55
Landkreis Ravensburg	17,58	21,12	0,88	19,33
Landkreis Sigmaringen	17,01	19,87	0,66	18,85
Mittelwert	17,30	20,48	0,81	18,94
Minimum	16,14	19,29	0,85	17,81
Maximum	18,54	21,15	0,63	19,83
Korrelation mit Verbesserung	-0,75	0,21	-	-0,27
Korrelation mit Mittelwert	0,72	0,74	-	-

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.7.: Fiktive fernere Lebenserwartung bei Geburt in ganz Baden-Württemberg

<b>Jahr</b>	<b>Deutsche Männer</b>	<b>Ausländische Männer</b>	<b>Deutsche Frauen</b>	<b>Ausländische Frauen</b>
1981	71,13	81,99	77,78	86,99
1982	71,44	81,09	78,14	87,57
1983	71,42	81,80	78,39	87,56
1984	72,49	83,13	79,11	89,43
1985	72,60	84,03	79,06	89,53
1986	72,87	83,90	79,55	89,87
1987	73,22	82,49	79,85	88,69
1988	73,43	81,21	80,16	85,80
1989	73,46	81,56	80,10	86,26
1990	73,57	80,18	80,00	86,71
1991	73,79	81,74	80,39	86,46
1992	74,31	82,28	80,67	87,38
1993	74,13	82,00	80,63	87,54
1994	74,66	81,97	81,04	88,01
1995	74,68	82,86	81,11	88,72
1996	75,02	83,68	81,13	88,72
1997	75,55	83,60	81,57	88,45
1998	75,74	83,87	81,81	87,98
1999	76,08	83,28	81,86	88,44
2000	76,57	83,90	82,44	88,78
2001	76,79	85,26	82,61	88,92
Mittelwert	73,95	82,66	80,35	87,99
Verbesserung %	7,96	3,99	6,20	2,22
MW 1981-1986	71,99	82,66	78,67	88,49
MW 1988-2001	74,84	82,67	81,11	87,73
Differenz	2,85	0,01	2,43	-0,77

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.8.: Simulation: Standardabweichung  $e_0$  über die Kreise, Männer

<b>Jahr</b>	<b>STABW wahr</b>	<b>MW STABW simuliert</b>	<b>95% STABW simuliert</b>	<b>MAX STABW simuliert</b>
1981	1,1485	0,7200	0,8767	1,0424
1982	0,9015	0,7248	0,9027	1,0912
1983	0,9857	0,7042	0,8588	1,1275
1984	1,0141	0,6837	0,8194	1,1143
1985	1,0968	0,6809	0,8264	1,0618
1986	1,0327	0,6627	0,8011	0,9946
1987	0,8068	0,6666	0,8135	1,0290
1988	0,7238	0,6500	0,7815	1,0086
1989	0,7783	0,6413	0,7759	0,9991
1990	0,8419	0,6344	0,7721	0,9814
1991	0,7305	0,6224	0,7562	0,9547
1992	0,9429	0,6033	0,7341	0,9022
1993	0,8403	0,6075	0,7358	0,8677
1994	0,8045	0,6090	0,7331	0,8676
1995	0,8770	0,5940	0,7181	0,8878
1996	0,5737	0,5837	0,7152	0,8709
1997	0,7176	0,5833	0,7041	0,8099
1998	0,8321	0,5845	0,7030	0,8548
1999	0,8897	0,5924	0,7134	0,8816
2000	0,8117	0,5761	0,7052	0,9344
2001	0,8143	0,5933	0,7338	0,8598
MW	0,8650	0,6342	0,7705	0,9591

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.9.: Simulation: Standardabweichung  $e_0$  über die Kreise, Frauen

<b>Jahr</b>	<b>STABW wahr</b>	<b>MW STABW simuliert</b>	<b>95% STABW simuliert</b>	<b>MAX STABW simuliert</b>
1981	0,8064	0,6808	0,8292	1,0525
1982	1,0153	0,6809	0,8261	1,1764
1983	1,0296	0,6591	0,7951	0,9338
1984	0,7940	0,6513	0,7891	0,9454
1985	0,8445	0,6565	0,8070	0,9897
1986	1,0172	0,6263	0,7521	0,9131
1987	0,9658	0,6270	0,7527	0,9842
1988	0,7479	0,5997	0,7222	0,8466
1989	0,7952	0,6123	0,7412	0,9967
1990	0,6980	0,6026	0,7343	0,8670
1991	0,7572	0,5843	0,7034	0,8491
1992	0,7057	0,5763	0,7002	0,8440
1993	0,8032	0,5789	0,6933	0,8625
1994	0,8367	0,5737	0,6949	0,9193
1995	0,7548	0,5742	0,6922	0,9173
1996	0,6135	0,5653	0,6932	0,7851
1997	0,7519	0,5497	0,6641	0,7566
1998	0,7446	0,5611	0,6837	0,8403
1999	0,7808	0,5619	0,6763	0,8255
2000	0,7296	0,5324	0,6550	0,7610
2001	0,7443	0,5440	0,6506	0,8135
MW	0,8065	0,5999	0,7265	0,8990

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.



A. Tabellen

Tabelle A.10.: Simulation: Standardabweichung  $e_{65}$  über die Kreise, Männer

<b>Jahr</b>	<b>STABW wahr</b>	<b>MW STABW simuliert</b>	<b>95% STABW simuliert</b>	<b>MAX STABW simuliert</b>
1981	0,4485	0,3172	0,3801	0,4585
1982	0,5338	0,3322	0,4039	0,4849
1983	0,4635	0,3390	0,4111	0,4966
1984	0,4965	0,3510	0,4170	0,5111
1985	0,5181	0,3392	0,4048	0,4695
1986	0,6017	0,3415	0,4078	0,4851
1987	0,4829	0,3458	0,4172	0,4828
1988	0,4781	0,3451	0,4120	0,5005
1989	0,4156	0,3447	0,4146	0,4919
1990	0,4430	0,3362	0,4034	0,4672
1991	0,5052	0,3363	0,4015	0,4530
1992	0,5783	0,3355	0,4064	0,4848
1993	0,4964	0,3278	0,3931	0,4704
1994	0,5051	0,3290	0,3919	0,4727
1995	0,4653	0,3224	0,3904	0,5229
1996	0,4498	0,3213	0,3846	0,5065
1997	0,4257	0,3214	0,3860	0,4646
1998	0,4598	0,3209	0,3912	0,4647
1999	0,4459	0,3173	0,3844	0,4708
2000	0,4318	0,3172	0,3847	0,4869
2001	0,4947	0,3167	0,3842	0,4498
MW	0,4828	0,3313	0,3986	0,4807

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.11.: Simulation: Standardabweichung  $e_{65}$  über die Kreise, Frauen

<b>Jahr</b>	<b>STABW wahr</b>	<b>MW STABW simuliert</b>	<b>95% STABW simuliert</b>	<b>MAX STABW simuliert</b>
1981	0,5308	0,2786	0,3361	0,3949
1982	0,6485	0,2906	0,3458	0,4019
1983	0,5334	0,2918	0,3507	0,4083
1984	0,5454	0,3013	0,3614	0,4157
1985	0,6194	0,2927	0,3540	0,4049
1986	0,6742	0,2941	0,3522	0,4391
1987	0,5809	0,2985	0,3568	0,4179
1988	0,5567	0,2970	0,3559	0,4204
1989	0,6009	0,2955	0,3554	0,4438
1990	0,5666	0,2867	0,3410	0,4176
1991	0,6134	0,2870	0,3460	0,4073
1992	0,5352	0,2915	0,3503	0,4430
1993	0,6401	0,2888	0,3463	0,3983
1994	0,5432	0,2917	0,3464	0,4311
1995	0,5016	0,2916	0,3507	0,4034
1996	0,4256	0,2892	0,3458	0,3987
1997	0,4476	0,2921	0,3509	0,3996
1998	0,4719	0,2942	0,3562	0,4679
1999	0,5050	0,2926	0,3484	0,4080
2000	0,4355	0,2928	0,3507	0,4003
2001	0,4031	0,2908	0,3482	0,3974
MW	0,5419	0,2919	0,3500	0,4152

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.12.: Simulation: Standardabweichung  $e_0$  über die Zeit, Frauen

Kreis	STABW	MW STABW	2,5% STABW	97,5% STABW
	wahr	simuliert	simuliert	simuliert
Stadtkreis Stuttgart	1,3107	1,4824	1,3217	1,6296
Landkreis Böblingen	1,2951	1,5043	1,3084	1,7083
Landkreis Esslingen	1,4714	1,4854	1,3312	1,6426
Landkreis Göppingen	1,4995	1,5294	1,3058	1,7509
Landkreis Ludwigsburg	1,6162	1,4858	1,3106	1,6495
Rems-Murr-Kreis	1,4723	1,4928	1,3201	1,6671
Stadtkreis Heilbronn	1,4248	1,6203	1,3057	1,9301
Landkreis Heilbronn	1,7217	1,5171	1,3133	1,7267
Hohenlohekreis	1,7472	1,6562	1,3026	2,0030
Landkreis Schwäbisch Hall	1,4223	1,5559	1,3016	1,8397
Main-Tauber-Kreis	1,6864	1,6077	1,3322	1,8861
Landkreis Heidenheim	1,5921	1,5855	1,3094	1,8685
Ostalbkreis	1,3436	1,5111	1,2968	1,7068
Stadtkreis Baden-Baden	2,5832	1,8506	1,3686	2,3547
Stadtkreis Karlsruhe	1,7712	1,5202	1,3099	1,7415
Landkreis Karlsruhe	1,5281	1,4960	1,3113	1,6696
Landkreis Rastatt	1,6357	1,5466	1,3106	1,7865
Stadtkreis Heidelberg	1,4082	1,6077	1,3075	1,9305
Stadtkreis Mannheim	1,4580	1,5123	1,2991	1,7244
Neckar-Odenwald-Kreis	1,4409	1,5820	1,3074	1,8853
Rhein-Neckar-Kreis	1,4977	1,4848	1,3331	1,6383
Stadtkreis Pforzheim	1,4215	1,6254	1,3023	1,9668
Landkreis Calw	1,5339	1,5664	1,2843	1,8455
Enzkreis	1,3976	1,5578	1,3100	1,8152
Landkreis Freudenstadt	1,5922	1,6299	1,3092	1,9646
Stadtkreis Freiburg	1,4588	1,5537	1,2978	1,8224
Landkreis Breisgau-Hochschwarzwald	1,2179	1,5356	1,2918	1,7678
Landkreis Emmendingen	1,3992	1,5770	1,2833	1,8633
Ortenaukreis	1,5799	1,4994	1,3118	1,6772
Landkreis Rottweil	1,4590	1,5926	1,3025	1,8719
Schwarzwald-Baar-Kreis	1,1935	1,5400	1,2942	1,7797
Landkreis Tuttlingen	1,9735	1,6082	1,3023	1,9006
Landkreis Konstanz	1,1852	1,5271	1,2933	1,7554
Landkreis Lörrach	1,2356	1,5452	1,3086	1,7959
Landkreis Waldshut	1,2048	1,5711	1,2935	1,8503
Landkreis Reutlingen	1,4797	1,5225	1,3139	1,7507
Landkreis Tübingen	1,6051	1,5486	1,2965	1,8303
Zollernalbkreis	1,8317	1,5510	1,2873	1,8023
Stadtkreis Ulm	1,4458	1,6390	1,3285	1,9643
Alb-Donau-Kreis	1,3636	1,5523	1,3002	1,8172
Landkreis Biberach	1,5116	1,5595	1,2937	1,8364
Bodenseekreis	1,3202	1,5584	1,3122	1,8051
Landkreis Ravensburg	1,5438	1,5278	1,3058	1,7471
Landkreis Sigmaringen	1,4606	1,6112	1,2879	1,9311
Mittelwert	1,5078	1,5599	1,3072	1,8159

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.13.: Simulation von Messfehlern: Standardabweichung  $e_{65}$  über die Kreise, Männer

<b>Jahr</b>	<b>STABW wahr</b>	<b>MW STABW simuliert</b>	<b>2,5% STABW simuliert</b>	<b>97,5% STABW simuliert</b>
1988	0,4781	0,4589	0,4576	0,4601
1989	0,4156	0,3843	0,3826	0,3860
1990	0,4430	0,4116	0,4091	0,4142
1991	0,5052	0,4679	0,4651	0,4707
1992	0,5783	0,5671	0,5639	0,5702
1993	0,4964	0,4981	0,4943	0,5020
1994	0,5051	0,5132	0,5093	0,5167
1995	0,4653	0,4554	0,4516	0,4594
1996	0,4498	0,4508	0,4460	0,4561
1997	0,4257	0,4427	0,4380	0,4475
1998	0,4598	0,4224	0,4176	0,4270
1999	0,4459	0,4349	0,4290	0,4408
2000	0,4318	0,4693	0,4624	0,4760
2001	0,4947	0,4896	0,4809	0,5000
MW	0,4711	0,4619	0,4577	0,4662

*Quelle:* Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.14.: Deskriptive Statistiken zu den verwendeten Variablen in Kapitel 5.

Variable	Zahl der Beobachtungen	Mittelwert	Standardabweichung	Minimum	Maximum
al	880	3,64	1,25	1,39	8,24
auslq_f	968	9,40	3,57	3,05	26,36
auslq_m	968	11,69	4,50	3,67	28,48
bett_reg	352	3468,75	3529,79	360,00	11623,00
co_98	156	1,16	0,37	0,05	2,40
e_alt1865_f	968	39,99	0,94	36,79	42,88
e_alt1865_m	968	39,41	1,14	35,57	42,44
e_alt66p_f	968	41,28	1,63	37,15	48,18
e_alt66p_m	968	37,46	1,48	33,44	43,63
e_alter_f	968	41,28	1,63	37,15	48,18
e_alter_m	968	37,46	1,48	33,44	43,63
ek_mw_125	572	1647,20	77,77	1417,39	1898,73
ek_mw_vdr_f	176	470,90	54,47	381,28	606,16
ek_mw_vdr_m	176	1044,13	72,70	890,82	1186,79
em	176	2,61	0,48	1,71	4,24
ewq_ao_f	352	45,80	12,11	29,42	82,09
ewq_ao_m	352	57,38	13,24	36,95	101,13
ewq_wo_f	176	43,64	3,37	34,23	52,58
ewq_wo_m	176	55,39	6,45	36,43	66,77
gebfrei	968	8,86	6,61	2,80	33,10
k_alt66p_f	968	2,07	0,08	1,86	2,35
k_alt66p_m	968	2,27	0,11	1,98	2,75
k_alter_f	968	2,07	0,08	1,86	2,35
k_alter_m	968	2,27	0,11	1,98	2,75
ma_ao	880	64,20	4,62	45,37	73,27
ma_ao_f	352	66,70	3,25	53,27	72,83
ma_ao_m	352	67,62	4,02	55,49	75,33
ma_wo_f	176	66,88	3,63	55,69	73,09
ma_wo_m	176	65,75	5,55	46,78	73,89
r_stz_1865f	968	194,30	33,99	105,46	351,45
r_stz_1865m	968	386,97	53,82	191,98	620,16
r_stz_66plus_f	968	4646,21	363,12	3259,52	5896,29
r_stz_66plus_m	968	5882,24	799,13	4003,87	8146,11
r_stz_f	968	1012,70	134,24	634,69	1691,57
r_stz_m	968	954,83	128,58	603,18	1638,57
sh_f	352	2,82	1,31	0,07	8,01
sh_m	352	2,28	1,22	0,03	6,84
sk_alt1865_f	968	0,12	0,10	-0,09	0,53

Fortsetzung auf der nächsten Seite.

Variable	Zahl der Beobachtungen	Mittelwert	Standard- abweichung	Minimum	Maximum
sk_alt1865_m	968	0,16	0,09	-0,08	0,55
sk_alt66p_f	968	0,17	0,08	-0,10	0,42
sk_alt66p_m	968	0,25	0,10	-0,04	0,57
sk_alter_f	968	0,17	0,08	-0,10	0,42
sk_alter_m	968	0,25	0,10	-0,04	0,57
so2_mw	148	0,01	0,00	0,00	0,01
stadt_dummy	968	0,20	0,40	0,00	1,00
std_alt1865_f	968	13,52	0,40	12,64	14,62
std_alt1865_m	968	13,25	0,32	12,29	14,12
std_alt66p_f	968	23,20	0,42	22,05	24,08
std_alt66p_m	968	21,14	0,52	19,54	22,56
std_alter_f	968	23,20	0,42	22,05	24,08
std_alter_m	968	21,14	0,52	19,54	22,56
uni_ao	880	5,93	3,39	1,77	20,11
uni_ao_f	352	4,53	2,35	1,95	13,42
uni_ao_m	352	9,72	4,96	3,79	25,62
uni_wo_f	176	5,70	3,81	2,43	22,42
uni_wo_m	176	11,56	5,55	5,44	32,62
verkehr	968	6,12	2,75	2,70	15,90
wald	968	36,68	11,79	11,70	62,60
wand_ab_66pf	968	1,40	0,70	0,62	10,79
wand_ab_66pm	968	1,28	0,72	0,53	11,39
wand_ab_f	968	4,76	2,07	2,31	25,06
wand_ab_m	968	5,95	2,63	2,64	26,70
wand_zu_66pf	968	1,61	0,76	0,65	10,51
wand_zu_66pm	968	1,43	0,77	0,59	10,55
wand_zu_f	968	5,37	2,20	2,45	24,19
wand_zu_m	968	6,58	2,74	2,65	25,63

*Quelle:* Statistisches Landesamt Baden-Württemberg, VDR, Sozialministerium Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.15.: Sterblichkeit deutscher Männer, Zeitreihe 1999-2002. Abhängige Variable  $r\_stz\_m$ , 176 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20***	-32,908	8,719
jahrd21***	-59,600	10,023
jahrd22***	-36,504	12,202
e_alter_m***	148,614	20,041
std_alter_m*	-47,240	25,504
sk_alter_m***	1162,062	322,713
ek_mw_vdr_m	-0,230	0,160
ma_wo_m	1,594	3,144
uni_wo_m**	-5,912	2,680
ewq_wo_m***	6,026	1,607
stadt_dummy	16,739	37,129
_cons***	-4126,203	820,446
$R^2_{gesamt}$	0,827	
$R^2_{between}$	0,889	
$R^2_{within}$	0,295	
$\sigma_\lambda$	33,420	
$\sigma_\varepsilon$	35,047	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,476	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, VDR, eigene Berechnungen.

Tabelle A.16.: Sterblichkeit deutscher Männer zwischen 18 und 65 Jahren, Zeitreihe 1999-2002. Abhängige Variable r\_stz\_1865\_m, 176 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20	3,748	8,147
jahrd21	3,944	10,736
jahrd22**	-30,914	12,985
e.alt1865_m***	52,553	18,867
std.alt1865_m	-21,706	26,521
sk.alt1865_m	315,408	192,193
ek_mw_vdr_m**	-0,212	0,088
ma_wo_m	1,733	1,952
uni_wo_m	-0,840	1,754
ewq_wo_m***	3,284	0,874
em**	16,593	7,847
al***	15,842	4,574
stadt_dummy***	51,304	14,345
_cons**	-1735,106	847,316
$R^2_{gesamt}$	0,630	
$R^2_{between}$	0,827	
$R^2_{within}$	0,273	
$\sigma_\lambda$	12,182	
$\sigma_\varepsilon$	29,398	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,147	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, VDR, eigene Berechnungen.



A. Tabellen

Tabelle A.17.: Alterssterblichkeit deutscher Männer, Zeitreihe 1999-2002. Abhängige Variable  $r\_stz\_66plus\_m$ , 176 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20***	-240,101	49,729
jahrd21***	-442,109	53,241
jahrd22**	-210,461	93,437
e_alt66p_m	120,420	334,118
std_alt66p_m	-15,055	383,924
sk_alt66p_m	-298,196	1022,075
ek_mw_vdr_m**	-1,709	0,780
ma_wo_m	17,492	19,488
uni_wo_m*	30,048	18,151
ewq_wo_m*	17,684	9,332
em*	168,421	88,371
stadt_dummy*	332,850	187,971
wand_zu_66pm	-25,899	123,321
wand_ab_66pm	32,412	134,715
_cons	-4720,207	23657,980
$R^2_{gesamt}$	0,395	
$R^2_{between}$	0,333	
$R^2_{within}$	0,478	
$\sigma_\lambda$	193,296	
$\sigma_\varepsilon$	198,230	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,487	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, VDR, eigene Berechnungen.

Tabelle A.18.: Sterblichkeit deutscher Frauen, Zeitreihe 1999-2002. Abhängige Variable  $r\_stz\_f$ , 176 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20*	-17,296	9,985
jahrd21***	-45,479	11,576
jahrd22	8,442	14,374
e_alter_f***	141,259	22,119
std_alter_f	-40,801	37,198
sk_alter_f***	1230,939	400,332
ek_mw_vdr_f***	-0,654	0,238
ma_wo_f	-4,851	3,197
uni_wo_f***	-15,398	5,640
ewq_wo_f**	7,073	3,205
stadt_dummy	49,107	47,383
_cons***	-3742,077	1248,328
$R^2_{gesamt}$	0,845	
$R^2_{between}$	0,897	
$R^2_{within}$	0,328	
$\sigma_\lambda$	43,817	
$\sigma_\varepsilon$	40,439	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,540	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, VDR, eigene Berechnungen.

A. Tabellen

Tabelle A.19.: Alterssterblichkeit deutscher Frauen, Zeitreihe 1999-2002. Abhängige Variable  $r\_stz\_66plus\_f$ , 176 Beobachtungen.

<b>Variable</b>	<b>Koeffizient</b>	<b>Standardabweichung</b>
jahrd20	-115,752	94,577
jahrd21**	-338,990	148,690
jahrd22	-129,966	155,746
e.alt66p_f	7,391	382,685
std.alt66p_f	398,222	590,454
sk.alt66p_f	-515,841	3119,961
k.alt66p_f	30,509	1161,710
ek_mw_vdr_f***	-2,942	1,082
ma_wo_f	0,359	9,817
uni_wo_f*	26,575	15,021
ewq_wo_f**	36,024	14,156
em**	173,905	77,103
stadt_dummy**	323,948	158,271
wand_zu_66pf***	291,393	84,006
wand_ab_66pf***	-268,237	94,484
_cons	176,410	25459,772
$R^2_{gesamt}$	0,436	
$R^2_{between}$	0,487	
$R^2_{within}$	0,374	
$\sigma_\lambda$	140,961	
$\sigma_\varepsilon$	183,191	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,372	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, VDR, eigene Berechnungen.

Tabelle A.20.: Sterblichkeit deutscher und ausländischer Männer, Zeitreihe 1983-1995.  
Abhängige Variable r\_stz\_m, 572 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd4***	-68,384	8,530
jahrd5***	-68,523	9,353
jahrd6***	-92,565	10,308
jahrd7***	-133,061	12,520
jahrd8***	-150,113	15,125
jahrd9***	-157,684	17,532
jahrd10***	-143,090	16,752
jahrd11***	-160,031	16,375
jahrd12***	-184,985	16,808
jahrd13***	-181,456	18,932
jahrd14***	-220,144	20,766
jahrd15***	-237,129	22,194
e_alter_m***	68,043	11,025
std_alter_m***	102,455	14,949
sk_alter_m	114,727	143,756
ek_mw_125**	-0,156	0,065
ewq_ao	-0,531	0,711
ma_ao	-1,101	1,562
uni_ao	-1,136	2,859
stadt_dummy**	80,493	37,497
auslq_m**	-3,630	1,635
_cons***	-3204,504	489,392
$R_{gesamt}^2$	0,839	
$R_{between}^2$	0,898	
$R_{within}^2$	0,620	
$\sigma_\lambda$	36,398	
$\sigma_\varepsilon$	37,919	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,480	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.21.: Alterssterblichkeit deutscher und ausländischer Männer, Zeitreihe 1983-1995. Abhängige Variable `r_stz_66plus_m`, 572 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd4***	-334,391	65,697
jahrd5***	-302,186	71,013
jahrd6***	-478,165	77,223
jahrd7***	-737,744	92,746
jahrd8***	-820,617	111,660
jahrd9***	-836,089	129,291
jahrd10***	-740,714	122,617
jahrd11***	-903,535	118,612
jahrd12***	-1026,080	119,961
jahrd13***	-1068,347	134,488
jahrd14***	-1320,022	147,322
jahrd15***	-1462,592	157,526
e_alt66p_m	67,682	77,281
std_alt66p_m	-64,021	106,550
sk_alt66p_m	-231,400	1025,482
ek_mw_125***	-1,512	0,485
ewq_ao	-5,676	4,808
ma_ao	-17,006	10,763
uni_ao	-22,149	19,708
stadt_dummy	221,021	251,957
auslq_m	-2,749	11,673
_cons***	9877,973	3412,018
$R^2_{\text{gesamt}}$	0,591	
$R^2_{\text{between}}$	0,229	
$R^2_{\text{within}}$	0,684	
$\sigma_\lambda$	219,959	
$\sigma_\varepsilon$	292,264	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,362	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.22.: Sterblichkeit deutscher und ausländischer Frauen, Zeitreihe 1983-1995.  
Abhängige Variable  $r\_stz\_f$ , 572 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd4***	-72,163	9,928
jahrd5***	-57,160	11,120
jahrd6***	-84,106	12,511
jahrd7***	-103,543	14,755
jahrd8***	-108,126	17,041
jahrd9***	-102,323	19,587
jahrd10***	-72,119	19,568
jahrd11***	-88,078	20,547
jahrd12***	-105,298	23,024
jahrd13***	-88,879	25,518
jahrd14***	-129,147	27,519
jahrd15***	-136,080	28,848
e_alter_f***	84,334	13,563
std_alter_f***	105,623	24,241
sk_alter_f**	451,455	202,186
ek_mw_125	-0,046	0,087
ewq_ao	-1,236	0,931
ma_ao	-0,267	1,938
uni_ao	-1,568	3,807
stadt_dummy	18,856	50,360
auslq_f*	-3,787	2,239
_cons***	-4685,161	846,330
$R_{gesamt}^2$	0,776	
$R_{between}^2$	0,853	
$R_{within}^2$	0,190	
$\sigma_\lambda$	49,094	
$\sigma_\varepsilon$	43,650	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,559	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.23.: Sterblichkeit deutscher Männer, Zeitreihe 1995-2002. Abhängige Variable  $r\_stz\_m$ , 352 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd16**	-19,411	8,264
jahrd17***	-62,481	9,873
jahrd18***	-74,029	9,636
jahrd19***	-89,665	10,787
jahrd20***	-113,301	13,377
jahrd21***	-136,108	15,083
jahrd22***	-115,452	14,499
e_alter_m***	114,598	15,333
std_alter_m	-46,430	31,489
sk_alter_m***	678,457	248,023
k_alter_m	-146,696	202,958
stadt_dummy	-48,452	39,131
wand_zu_m	-3,824	3,839
wand_ab_m	4,304	4,013
gebfrei*	6,448	3,786
verkehr*	-15,721	9,434
wald	0,273	0,718
ma_ao_m**	5,866	2,304
uni_ao_m	-4,191	2,641
ewq_ao_m**	1,504	0,646
al	4,232	5,837
sh_m**	11,454	5,592
bett_reg**	-0,005	0,003
_cons***	-2651,730	1023,046
$R^2_{gesamt}$	0,872	
$R^2_{between}$	0,927	
$R^2_{within}$	0,548	
$\sigma_\lambda$	31,126	
$\sigma_\varepsilon$	34,240	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,452	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

Tabelle A.24.: Sterblichkeit deutscher Frauen, Zeitreihe 1995-2002. Abhängige Variable  $r_{stz\_f}$ , 352 Beobachtungen.

<b>Variable</b>	<b>Koeffizient</b>	<b>Standardabweichung</b>
jahrd16	11,733	10,519
jahrd17	-6,349	13,326
jahrd18***	-41,782	12,180
jahrd19***	-45,299	13,179
jahrd20***	-60,716	15,715
jahrd21***	-88,414	17,804
jahrd22	-21,151	16,817
e_alter_f***	98,038	18,158
std_alter_f	-3,810	42,485
sk_alter_f	390,251	295,459
k_alter_f	61,506	307,552
stadt_dummy	29,182	70,074
wand_zu_f	2,945	4,678
wand_ab_f	-1,124	4,774
gebfrei	-0,101	5,017
verkehr	2,415	12,197
wald	1,084	1,023
ma_ao_f	-1,824	2,720
uni_ao_f**	-13,545	6,209
ewq_ao_f	-0,554	1,456
al	-5,254	7,311
sh_f*	10,529	6,202
bett_reg*	-0,007	0,004
_cons*	-2994,633	1650,127
$R^2_{gesamt}$	0,848	
$R^2_{between}$	0,908	
$R^2_{within}$	0,335	
$\sigma_\lambda$	45,395	
$\sigma_\varepsilon$	43,267	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,524	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, Sozialministerium Baden-Württemberg, eigene Berechnungen.



A. Tabellen

Tabelle A.25.: Alterssterblichkeit deutscher Frauen, Zeitreihe 1995-2002. Abhängige Variable  $r\_stz\_f$ , 352 Beobachtungen.

<b>Variable</b>	<b>Koeffizient</b>	<b>Standardabweichung</b>
jahrd16	60,008	52,406
jahrd17	-15,105	75,310
jahrd18	-104,929	82,622
jahrd19	-128,677	88,702
jahrd20	-182,548	118,254
jahrd21**	-358,453	146,203
jahrd22	-49,939	123,424
e_alt66p_f	106,632	262,007
std_alt66p_f	38,040	391,158
sk_alt66p_f	-1378,015	1796,877
k_alt66p_f	299,071	650,629
stadt_dummy	14,930	233,651
wand_zu_66pf**	101,453	40,691
wand_ab_66pf**	-85,362	40,825
gebfrei	24,990	16,359
verkehr	-28,554	40,112
wald	5,001	3,348
ma_ao_f_q	4,482	8,866
uni_ao_f_q	-24,478	17,496
ewq_ao_f	-3,452	4,730
al	6,474	27,735
sh_f	22,848	24,954
bett_reg	-0,009	0,011
_cons	-4088,670	17538,529
$R^2_{gesamt}$	0,370	
$R^2_{between}$	0,415	
$R^2_{within}$	0,327	
$\sigma_\lambda$	129,774	
$\sigma_\varepsilon$	186,856	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,325	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, Sozialministerium Baden-Württemberg, eigene Berechnungen.

Tabelle A.26.: Sterblichkeit deutscher Männer, Zeitreihe 1983-2002. Abhängige Variable  $r\_stz\_m$ , 880 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd4***	-67,171	9,325
jahrd5***	-65,139	9,934
jahrd6***	-90,890	10,892
jahrd7***	-127,378	11,998
jahrd8***	-134,892	13,351
jahrd9***	-122,774	15,947
jahrd10***	-111,823	17,711
jahrd11***	-141,121	18,861
jahrd12***	-179,206	20,752
jahrd13***	-162,390	22,711
jahrd14***	-198,585	23,938
jahrd15***	-205,820	24,922
jahrd16***	-225,125	26,058
jahrd17***	-273,718	26,756
jahrd18***	-292,866	27,139
jahrd19***	-322,398	27,590
jahrd20***	-362,369	28,326
jahrd21***	-397,290	28,852
jahrd22***	-365,303	29,083
e_alter_m***	89,818	10,565
std_alter_m***	62,052	18,908
sk_alter_m**	365,083	162,875
k_alter_m	-101,164	95,199
stadt_dummy	32,629	38,407
wand_zu_m**	-6,834	2,736
wand_ab_m*	4,735	2,775
gebfrei	-2,131	2,709
verkehr	9,621	6,038
wald**	1,523	0,720
ma_ao***	-3,373	1,276
uni_ao**	-4,704	2,208
ewq_ao	0,655	0,625
_cons***	-3260,200	511,498
$R^2_{gesamt}$	0,833	
$R^2_{between}$	0,879	
$R^2_{within}$	0,699	
$\sigma_\lambda$	39,568	
$\sigma_\varepsilon$	41,545	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,476	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.27.: Alterssterblichkeit deutscher Frauen, Zeitreihe 1983-2002. Kovariante Wanderung, Flächennutzung, Bildung. Abhängige Variable r\_stz\_66pluf, 880 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd4***	-361,46411	67,94782999
jahrd5**	-285,8785	123,2954581
jahrd6***	-455,60776	151,4176773
jahrd7***	-659,24709	156,7464481
jahrd8***	-744,15265	156,3467684
jahrd9***	-805,16036	154,4466963
jahrd10***	-725,64122	155,7716373
jahrd11***	-819,60095	155,2178018
jahrd12***	-927,9219	156,3216236
jahrd13***	-867,73611	158,4033049
jahrd14***	-1038,1642	159,3854667
jahrd15***	-1043,008	161,542626
jahrd16***	-996,23583	162,78722
jahrd17***	-1083,3885	162,4829222
jahrd18***	-1181,6843	162,8173517
jahrd19***	-1180,2475	166,0122706
jahrd20***	-1200,1793	173,3873294
jahrd21***	-1357,6721	180,5401673
jahrd22***	-1001,3516	179,0243447
e_alt66p_f***	481,28743	125,999
std_alt66p_f	73,893384	179,0140302
sk_alt66p_f	996,61463	872,0004186
k_alt66p_f***	-759,37044	281,4944547
stadt_dummy*	-275,04213	158,4415539
wand_zu_66pf***	152,48515	27,72251864
wand_ab_66pf***	-146,33852	26,50193408
gebfrei	2,2318303	12,2100258
verkehr*	45,552624	27,15861153
wald*	4,9606315	2,794143411
ma_ao	-7,5530314	5,766136922
uni_ao**	-22,315427	9,566368695
ewq_ao	-1,6831069	2,838661216
_cons***	-29764,322	8545,718928
<hr/>		
R <sup>2</sup> <sub>gesamt</sub>	0,33911041	
R <sup>2</sup> <sub>between</sub>	0,24492548	
R <sup>2</sup> <sub>within</sub>	0,41007698	
$\sigma_\lambda$	148,93207	
$\sigma_\varepsilon$	211,23006	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,3320528	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Tabelle A.28.: Sterblichkeit deutscher und ausländischer Männer, Zeitreihe 1999-2002.  
Abhängige Variable r\_stz\_m, 148 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20***	-40,520172	9,991800488
jahrd21***	-77,723376	11,23851503
jahrd22***	-48,72355	14,19697644
e_alter_m***	61,263012	19,08573053
std_alter_m***	71,191332	19,95276071
sk_alter_m	95,689528	312,2793493
auslq_m***	-8,3533626	2,277078787
stadt_dummy**	99,835585	42,17341698
so2_mw	4296,5457	3594,491619
co_98*	-27,692	14,24619072
_cons***	-2921,1061	837,1713564
$R^2_{gesamt}$	0,75995763	
$R^2_{between}$	0,85266768	
$R^2_{within}$	0,22125713	
$\sigma_\lambda$	33,044155	
$\sigma_\varepsilon$	32,546972	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,5075796	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

A. Tabellen

Tabelle A.29.: Sterblichkeit deutscher und ausländischer Frauen, Zeitreihe 1999-2002.  
Abhängige Variable  $r_{stz\_f}$ , 148 Beobachtungen.

Variable	Koeffizient	Standardabweichung
jahrd20	-9,1571199	11,78807194
jahrd21***	-41,567308	12,93728256
jahrd22	12,186882	16,30097083
e.alter_f***	70,802865	17,09194693
std.alter_f**	60,850136	25,07003371
sk.alter_f	153,94465	277,509117
auslq_f***	-11,906249	3,132474294
stadt_dummy*	80,830022	47,80233781
so2_mw	4026,1496	4239,973703
co_98	21,421226	16,70695304
_cons***	-3373,8124	1000,168586
$R^2_{gesamt}$	0,79997001	
$R^2_{between}$	0,87811933	
$R^2_{within}$	0,27960413	
$\sigma_\lambda$	37,79231	
$\sigma_\varepsilon$	37,216967	
$\frac{\sigma_\lambda}{\sigma_\lambda + \sigma_\varepsilon}$	0,50766983	

\* signifikant auf dem 10%-Niveau, \*\* 5%-Niveau, \*\*\* 1%-Niveau.

Quelle: Statistisches Landesamt Baden-Württemberg, eigene Berechnungen.

# Abkürzungsverzeichnis

$a_x$	Durchschnittliche Zahl der zwischen $x$ und $x + 1$ verlebten Lebensjahre der in dieser Altersklasse gestorbenen Personen.
al	Arbeitslose, Quote der Personen zwischen 18 und 65 Jahren.
alter	Alter in Jahren.
auslq_f	Ausländerquote, Frauen.
auslq_m	Ausländerquote, Männer.
bett_reg	Zahl der aufgestellten Betten pro „Region“ (s. Abschnitt 5.1).
bett1000	Aufgestellte Krankenhausbetten auf 1000 Einwohner.
co_98	98%-Quantil der Kohlenmonoxidimmissionen.
$e_{65}$	fiktive fernere Lebenserwartung im Alter 65.
$e_0$	fiktive Lebenserwartung bei Geburt.
e.alt1865_m	Erwartungswert der Altersverteilung zwischen 18 und 65 Jahren, Männer.
e.alt66p_f	Erwartungswert der Altersverteilung über 65 Jahren, Frauen.
e.alt66p_m	Erwartungswert der Altersverteilung über 65 Jahren, Männer.
e.alter_f	Erwartungswert der Altersverteilung, Frauen.
e.alter_m	Erwartungswert der Altersverteilung, Männer.
ek_mw_125	Mittelwert der Einkommen auf Basis der Einkommensteuerstatistik, obere Grenze der Einkommen 125.000 DM.
ek_mw_vdr_f	Mittelwert der Einkommen auf Basis der Rentenzahlungen, Frauen.
ek_mw_vdr_m	Mittelwert der Einkommen auf Basis der Rentenzahlungen, Männer.
EkSt	Einkommensteuer.
em	Quote der Bezieher von Erwerbsminderungsrenten an der Bevölkerung zwischen 18 und 65 Jahren.
ewq_ao	Erwerbstätigenquote am Arbeitsort, gesamt.
ewq_ao_f	Erwerbstätigenquote am Arbeitsort, Frauen.

## Abkürzungsverzeichnis

ewq_ao_m	Erwerbstätigenquote am Arbeitsort, Männer.
ewq_wo_f	Erwerbstätigenquote am Wohnort, Frauen.
ewq_wo_m	Erwerbstätigenquote am Wohnort, Männer.
gebfrei	Anteil der Gebäude- und Freifläche an der Gesamtfläche.
$j$	Indikator für die Periode.
$k$	Indikator für den Kreis.
k.alt66p_f	Kurtosis der Altersverteilung über 65 Jahren, Frauen.
k.alt66p_m	Kurtosis der Altersverteilung über 65 Jahren, Männer.
k.alter_f	Kurtosis der Altersverteilung, Frauen.
k.alter_m	Kurtosis der Altersverteilung, Männer.
$L_x$	Erwartete Zahl der Lebensjahre in der Altersklasse $x$ , bedingt auf das Erreichen des $x$ . Geburtstags.
$m_x$	Mortalitätsrate in der Altersklasse $x$ .
ma_ao	Quote der Beschäftigten mit abgeschlossener Berufsausbildung am Arbeitsort, gesamt.
ma_ao_f	Quote der Beschäftigten mit abgeschlossener Berufsausbildung am Arbeitsort, Frauen.
ma_ao_m	Quote der Beschäftigten mit abgeschlossener Berufsausbildung am Arbeitsort, Männer.
ma_wo_f	Quote der Beschäftigten mit abgeschlossener Berufsausbildung am Wohnort, Frauen.
ma_wo_m	Quote der Beschäftigten mit abgeschlossener Berufsausbildung am Wohnort, Männer.
$n$	Indikator für die Nationalität.
o3_98	98%-Quantil der Ozonbelastung.
$p_x$	Bedingte Überlebenswahrscheinlichkeit von der Altersklasse $x$ zur Altersklasse $x + 1$ .
PM <sub>10</sub>	Schwebstäube mit einem aerodynamischen Durchmesser von weniger als $10\mu\text{g}/\text{m}^3$ (PM steht für Particulate Matter).
r_stz_1865m	Rohe Sterbeziffer der Bevölkerung von 18 bis 65 Jahren, Männer.
r_stz_66plus_f	Rohe Sterbeziffer der Bevölkerung über 65 Jahren, Frauen.
r_stz_66plus_m	Rohe Sterbeziffer der Bevölkerung über 65 Jahren, Männer.
r_stzf	Rohe Sterbeziffer, Frauen.

r_stzm	Rohe Sterbeziffer, Männer.
s	Indikator für das Geschlecht.
sh_f	Quote der Sozialhilfeempfänger an der Gesamtbevölkerung, Frauen.
sh_m	Quote der Sozialhilfeempfänger an der Gesamtbevölkerung, Männer.
sk_alt1865_m	Schiefe der Altersverteilung zwischen 18 und 65 Jahren, Männer.
sk_alt66p_f	Schiefe der Altersverteilung über 65 Jahren, Frauen.
sk_alt66p_m	Schiefe der Altersverteilung über 65 Jahren, Männer.
sk_alter_f	Schiefe der Altersverteilung, Frauen.
sk_alter_m	Schiefe der Altersverteilung, Männer.
so2_mw	Mittelwert der Schwefeldioxidimmissionen.
stadt_dummy	Dummy für die Stadtkreise.
std_alt1865_m	Standardabweichung der Altersverteilung zwischen 18 und 65 Jahren, Männer.
std_alt66p_f	Standardabweichung der Altersverteilung über 65 Jahren, Frauen.
std_alt66p_m	Standardabweichung der Altersverteilung über 65 Jahren, Männer.
std_alter_f	Standardabweichung der Altersverteilung, Frauen.
std_alter_m	Standardabweichung der Altersverteilung, Männer.
uni_ao	Quote der Beschäftigten mit höherem Bildungsabschluss am Arbeitsort, gesamt.
uni_ao_f	Quote der Beschäftigten mit höherem Bildungsabschluss am Arbeitsort, Frauen.
uni_ao_m	Quote der Beschäftigten mit höherem Bildungsabschluss am Arbeitsort, Männer.
uni_wo_f	Quote der Beschäftigten mit höherem Bildungsabschluss am Wohnort, Frauen.
uni_wo_m	Quote der Beschäftigten mit höherem Bildungsabschluss am Wohnort, Männer.
VDR	Verband der Deutschen Rentenversicherungsträger e. V..
verkehr	Anteil der Verkehrsfläche an der Gesamtfläche.
wald	Anteil von Wäldern an der Gesamtfläche.
wand_ab_66pf	Abwanderungsquote über 65 Jahren, Frauen.
wand_ab_66pm	Abwanderungsquote über 65 Jahren, Männer.
wand_ab_f	Abwanderungsquote, Frauen.



## *Abkürzungsverzeichnis*

wand_ab_m	Abwanderungsquote, Männer.
wand_zu_66pf	Zuwanderungsquote über 65 Jahren, Frauen.
wand_zu_66pm	Zuwanderungsquote über 65 Jahren, Männer.
wand_zu_f	Zuwanderungsquote, Frauen.
wand_zu_m	Zuwanderungsquote, Männer.
<i>x</i>	Indikator für die Altersklasse: Alter am letzten Geburtstag.

# Literaturverzeichnis

- ACHEN, C. H., UND W. P. SHIVELY (1995): *Cross-Level Inference*. University of Chicago Press, Chicago.
- ANSON, J. (2003): "Sex Differences in Mortality at the Local Level: An Analysis of Belgian Municipalities," *European Journal of Population*, 19, 1–28.
- CISCHINSKY, H. (2000): "Vergleiche gesundheitlicher Outcomes in Baden-Württemberg," Diplomarbeit, Universität Mannheim.
- COALE, A. J., P. DEMENY, UND B. VAUGHEN (1983): *Regional Model Life Tables and Stable Populations*. Academic Press, New York, 2 edn.
- CONGDON, P. (1995): "Life Table Analysis for Areas Using Vital Register Data," *European Journal of Population*, 11, 343–369.
- DINKEL, R. H. (1985): "The Seeming Paradox of Increasing Mortality in a Highly Industrialized Nation: The Example of the Soviet Union," *Population Studies*, 39, 87–89.
- DOMINICI, F., ET AL. (2003): "Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities," *American Journal of Epidemiology*, 157(12), 1055–1065.
- DORBRITZ, J., UND K. GÄRTNER (1995): "Bericht 1995 über die demographische Lage in Deutschland," *Zeitschrift für Bevölkerungswissenschaft*, 20(4), 339–448.
- DOW, W. H., K. A. GONZALEZ, UND L. ROSERO-BIXBY (2003): "Aggregation and Insurance-Mortality Estimation," NBER Working Paper Nr. 9827.
- DURKHEIM, E. (1897): *Le Suicide*. F. Alcan, Paris.
- EIS, D. (1997): "Welchen Einfluss hat die Umwelt?," in *Das Public Health Buch*, hrsg. v. F. W. Schwartz, B. Badura, R. Leidl, H. Raspe, und J. Siegrist, S. 51–80. Urban & Fischer, München, Jena.

## Literaturverzeichnis

- ESENWEIN-ROTHE, I. (1982): *Einführung in die Demographie*, Nr. 10 in *Statistische Studien*. Franz Steiner, Wiesbaden.
- FIREBAUGH, G. (1978): "A rule for inferring individual-level relationships from aggregate data," *American Sociological Review*, 43, 557–572.
- FUCHS, V. R. (1974): "Some Economic Aspects of Mortality in Developed Countries," in *The Economics of Health and Medical Care*, hrsg. v. M. Perlman, S. 174–193. London.
- GÄRTNER, K. (1990): "Sterblichkeit nach dem Familienstand," *Zeitschrift für Bevölkerungswissenschaft*, 16(1), 53–66.
- GREENLAND, S. (2001): "Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects," *International Journal of Epidemiology*, 30, 1343–1350.
- (2002): "A review of multilevel theory for ecologic analyses," *Statistics in Medicine*, 21, 389–395.
- GREENLAND, S., UND H. MORGENSTERN (1989): "Ecological bias, confounding, and effect modification," *International Journal of Epidemiology*, 18, 269–274.
- GREENLAND, S., UND J. ROBINS (1994): "Ecologic studies - biases, misconceptions, and counterexamples," *American Journal of Epidemiology*, 139, 747–760.
- GRÖNER, G. (2002): "Zu Entwicklung und regionalen Unterschieden der Sterblichkeit in Baden-Württemberg um 1990," in *Regionale Sterblichkeit in Deutschland*, hrsg. v. J. Cromm, und R. D. Scholz, S. 39–64. WiSoMed, Göttingen, Augsburg.
- GUTHRIE, K. A., UND L. SHEPPARD (2001): "Overcoming biases and misconceptions in ecological studies," *Journal of the Royal Statistical Society, Series A*, 164(1), 141–154.
- HENKE, W., UND H. MÜLLER (2002): "Regionale Mortalität in Rheinland-Pfalz unter besonderer Berücksichtigung soziostruktureller Indikatoren," in *Regionale Sterblichkeit in Deutschland*, hrsg. v. J. Cromm, und R. D. Scholz, S. 205–230. WiSoMed, Göttingen, Augsburg.
- HUMMER, R., R. ROGERS, UND I. EBERSTEIN (1998): "Sociodemographic Differentials in Adult Mortality: A Review of Analytic Approaches," *Population and Development Review*, 24(3), 553–578.
- HURD, M. D., ET AL. (2003): "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status," *Journal of Econometrics*, 112(1), 3–56.

- KITAGAWA, E., UND P. HAUSER (1972): *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Harvard University Press, Cambridge.
- KLEIN, T. (1993): "Soziale Determinanten der Lebenserwartung," *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 45, 712–730.
- (1998): "Der Heimeintritt alter Menschen und Chancen seiner Vermeidung," *Zeitschrift für Gerontologie und Geriatrie*, 31, 407–416.
- (2001): "Der Einfluss differentieller Mortalität auf den ökonomischen Status im Alter," in *Soziale Sicherung in einer dynamischen Gesellschaft. Festschrift für Richard Hauser*, hrsg. v. I. Becker, N. Ott, und G. Rolf, S. 486–502. Campus-Verlag, Frankfurt am Main.
- KLEIN, T., S. SCHNEIDER, UND H. LÖWEL (2001): "Bildung und Mortalität. Die Bedeutung gesundheitsrelevanter Aspekte des Lebensstils," *Zeitschrift für Soziologie*, 30(5), 384–400.
- LASSERRE, V., C. GUIHENNEUC-JOUYAUX, UND S. RICHARDSON (2000): "Biases in ecological studies: utility of including within-area distribution of confounders," *Statistics in Medicine*, 19, 45–59.
- LLERAS-MUNEY, A. (2002): "The Relationship Between Education and Adult Mortality in the United States," NBER Working Paper Nr. 8986.
- MUELLER, U. (2000): "Die Maßzahlen der Bevölkerungsstatistik," in *Handbuch der Demographie*, hrsg. v. U. Mueller, A. Dieckmann, und B. Nauck, S. 1–91. Springer, Berlin, Heidelberg, New York.
- NEUHAUS, J., UND J. D. KALBFLEISCH (1998): "Between- and within-cluster covariate effects in the analysis of clustered data," *Biometrics*, 54, 638–645.
- PLUMMER, M., UND D. CLAYTON (1996): "Estimation of Population Exposure in Ecological Studies," *Journal of the Royal Statistical Society, Series B*, 58(1), 113–126.
- PRENTICE, R. L., UND L. SHEPPARD (1995): "Aggregate Data Studies of Disease Risk Factors," *Biometrika*, 82(1), 113–125.
- PRESTON, S. H., P. HEUVELINE, UND M. GUILLOT (2000): *Demography - measuring and modelling population processes*. Blackwell Publishers, Oxford, Malden, MA.
- REIL-HELD, A. (2002): "Die Rolle intergenerationaler Transfers in Einkommen und Vermögen älterer Menschen in Deutschland," Dissertation, Universität Mannheim.

Literaturverzeichnis

- RICHARDSON, S., I. STUCKER, UND H. HEMON (1987): "Comparison of relative risks obtained in ecological and individual studies: some methodological considerations," *International Journal of Epidemiology*, 16, 111–120.
- ROBINSON, W. S. (1950): "Ecological correlations and the behavior of individuals," *American Sociological Review*, 15, 351–357.
- ROSENBAUM, P. R., UND D. B. RUBIN (1984): "Difficulties with regression analyses of age-adjusted rates," *Biometrics*, 40, 437–443.
- SAMET, J. M., ET AL. (2000): "Fine Particulate Air Pollution and Mortality in 20 U.S. Cities, 1987-1994," *New England Journal of Medicine*, 343(24), 1742–1749.
- SCHOLZ, R. D. (2002): "Zu methodischen Problemen und Grenzen der regionalen Sterblichkeitsmessung," in *Regionale Sterblichkeit in Deutschland*, hrsg. v. J. Cromm, und R. D. Scholz, S. 7–17. WiSoMed, Göttingen, Augsburg.
- SHEPPARD, L. (2003): "Insights on bias and information in group-level studies," *Biostatistics*, 4(2), 265–278.
- SMITH, D. P. (1992): *Formal Demography*, The Plenum Series on Demographic Methods and Population Analysis. Plenum Press, New York, London.
- SMITH, J. P. (1999): "Healthy Bodies and Thick Wallets: The Dual Relation Between Health and Socioeconomic Status," *Journal of Economic Perspectives*, 13(2), 145–166.
- SNYDER, S. E., UND W. N. EVANS (2002): "The Impact of Income on Mortality: Evidence from the Social Security Notch," NBER Working Paper Nr. 9197.
- SOZIALMINISTERIUM BADEN-WÜRTTEMBERG (2003): "Homepage," [www.sozialministerium.baden-wuerttemberg.de](http://www.sozialministerium.baden-wuerttemberg.de).
- STATISTISCHES BUNDESAMT (2003): "Verbraucherpreisindex und Index der Einzelhandelspreise. Lange Reihe 1948-2002," Wiesbaden.
- STATISTISCHES LANDESAMT BADEN-WÜRTTEMBERG (2000): "Leibing: Verteilung von Finanzmitteln auf Basis unzutreffender Einwohnerzahlen?," Eildienst, herausgegeben vom Statistischen Landesamt Baden-Württemberg.
- (2003): "Regionaldatenbank im Internet," [www.statistik.bwl.de](http://www.statistik.bwl.de).

- THEURL, E. (1991): "Regionale Unterschiede der Mortalität in Österreich - Ökonomische Relevanz, Bestandsaufnahme und ökonomische Ursachen," in *Bevölkerungsentwicklung und öffentliche Haushalte*, hrsg. v. S. Fickl, S. 169–203. Frankfurt am Main, New York.
- VAUPEL, J. W., K. G. MANTON, UND E. STALLARD (1979): "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality," *Demography*, 16(3), 439–454.
- VAUPEL, J. W., UND A. I. YASHIN (1985): "Heterogeneity's Ruse: Some Surprising Effects of Selection on Population Dynamics," *The American Statistician*, 39(3), 176–185.
- WILKINSON, R. G. (1996): *Unhealthy Societies: The Afflictions of Inequality*. Routledge, London.
- WILLE, E., UND W. RIED (1996): "Die Umwelt als Determinante des Gesundheitszustands," in *Ökologie und Marktwirtschaft - Probleme, Ursachen und Lösungen*, hrsg. v. P. Eichhorn, S. 165–188. Wiesbaden.
- WITTWER-BACKOFEN, U. (1999): *Disparitäten der Alterssterblichkeit im regionalen Vergleich*, Nr. 95 in Materialien zur Bevölkerungswissenschaft. Bundesinstitut für Bevölkerungsforschung, Wiesbaden.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, London.

## Discussion Paper Series

Mannheim Research Institute for the Economics of Aging Universität Mannheim

**To order copies, please direct your request to the author of the title in question.**

<b>Nr.</b>	<b>Autoren</b>	<b>Titel</b>	<b>Jahr</b>
33-03	Florian Heiss	Wie groß soll die Schwankungsreserve der gesetzlichen Rentenversicherung sein?	03
34-03	Axel Börsch-Supan, Christina B. Wilke	The German Public Pension System: How it Was, How it Will Be	03
35-03	Lothar Essig, Anette Reil-Held	Chancen und Risiken der „Riester-Rente“	03
36-03	Barbara Berkel Axel Börsch-Supan	Pension Reform in Germany: The Impact on Retirement Decisions	03
37-03	Axel Börsch-Supan, Anette Reil-Held and Christina Benita Wilke	How to make a Defined Benefit System Sustainable: The “Sustainability Factor” in the German Benefit Indexation Formula	03
38-03	Melanie Lührmann	Demographic Change, Foresight and International Capital Flows	03
39-03	Joachim Winter and Lothar Essig	Item nonresponse to financial questions in household surveys: An experimental study of interviewer and mode effects	03
40-03	Christoph Buchheim	Das Zusammenspiel von Wirtschaft, Bevölkerung und Wohlstand aus historischer Sicht	03
41-03	Oliver Lipps	Stochastische Bevölkerungsprognose für West- und Ostdeutschland	03
42-03	Axel Börsch-Supan	What are NDC Pension Systems? What Do They Bring to Reform Strategies?	03
43-04	Matthias Weiss	Employment of Skill Biased Technological Change when Benefits are Linked to Per-Capita Income	04
44-04	Alexander Ludwig and Torsten Sløk	The relationship between stock prices, house prices and consumption in OECD countries	04
45-04	Matthias Weiss	Skill-Biased Technological Change: Is there Hope for the Unskilled?	04
46-04	Hans-Martin von Gaudecker	Regionale Mortalitätsunterschiede in Baden- Württemberg	04