

Die Bevölkerungsentwicklung der Metropolregion Rhein-Neckar: Eine stochastische Bevölkerungsprognose auf Basis des Paradigmas funktionaler Daten

Philipp Deschermeier

Zusammenfassung: Vor dem Hintergrund des demographischen Wandels erfreuen sich stochastische Bevölkerungsprognosen wachsender Beliebtheit, denn sie liefern im Vergleich zu der für Vorausberechnungen häufig verwendeten Szenariotechnik eine wichtige Zusatzinformation: die prognostizierte Bevölkerung liegt innerhalb eines Schwankungsbereichs, dem eine Eintrittswahrscheinlichkeit zugeordnet werden kann. Allerdings erfordert dieser Ansatz lange Zeitreihen und detaillierte Informationen über die Determinanten der Bevölkerungsentwicklung (Fertilität, Mortalität und Nettomigration), die auf regionaler Ebene häufig nicht in ausreichender Tiefe verfügbar, sondern in der Regel zu Altersgruppen zusammengefasst sind. Stochastische Bevölkerungsprognosen beschränken sich deshalb meist auf die nationale Ebene. Jedoch ermöglichen es Methoden der funktionalen Datenanalyse, die erforderlichen demographischen Merkmale auf Altersjahre zu disaggregieren und als Datenbasis einer stochastischen Modellierung, auch auf regionaler Ebene, zu nutzen. In diesem Aufsatz werden am Beispiel der Bevölkerungsentwicklung der Metropolregion Rhein-Neckar dieser Ansatz und darauf aufbauende Modelle vorgestellt.

Schlagwörter: Metropolregion Rhein-Neckar · Stochastische
Bevölkerungsprognosen · Funktionale Datenanalyse

1 Einleitung

Der demographische Wandel verändert die Gesellschaft tief greifend: Das Verhältnis von Jungen zu Alten und von Erwerbstätigen zu Rentnern verschiebt sich zu Gunsten der älteren Personen. Sowohl die öffentliche Debatte als auch die wissenschaftliche Forschung in Europa konzentrieren sich weitgehend auf sozialpolitische Konsequenzen des demographischen Wandels. Insbesondere liegt der Fokus auf dem Finanzierungsproblem der staatlichen Rentenversicherung, wodurch der

negative Klang des Begriffs „Bevölkerungsalterung“ entstand. Dabei wird häufig die positive Seite des demographischen Wandels vergessen: Menschen haben eine höhere Lebenserwartung und eine längere aktive Lebensspanne (*Schnabel et al.* 2005: 3).

Die steigende Lebenserwartung, insbesondere die überdurchschnittlich sinkende Mortalität älterer Menschen, bewirkt bei anhaltend niedriger Geburtenhäufigkeit weit reichende Veränderungen der Altersstruktur. Obgleich auf nationaler Ebene nur ein geringfügiger Rückgang der Bevölkerung in den kommenden 20 Jahren prognostiziert ist (von 81 Mio. auf ca. 77 bis 79 Mio. Einwohner), wird die Anzahl der Menschen im erwerbsfähigen Alter deutlich zurückgehen (*Statistisches Bundesamt* 2009; *Fuchs/Dörfler* 2005a/b; *Börsch-Supan/Wilke* 2009). So steht die Gesellschaft in den kommenden Jahrzehnten vor einem tief greifenden makroökonomischen Strukturwandel, der alle wichtigen Märkte betrifft: Dem Arbeitsmarkt fehlen junge Erwerbstätige, die Produktmärkte müssen sich auf strukturell veränderte Konsumentenwünsche einstellen und auf dem Kapitalmarkt ändern sich Sparverhalten und die Nachfrage nach Vermögensanlagen (*Börsch-Supan* 2007). Auf regionaler Ebene divergiert die zukünftige Bevölkerungsentwicklung nach Zahl und Struktur aufgrund unterschiedlicher ökonomischer Bedingungen in hohem Maße, so dass sich aus den von Region zu Region abweichenden Konsequenzen für die Güternachfrage, den Wohnungsmarkt oder das Arbeitskräfteangebot ein steigender Bedarf an regionalen Bevölkerungsprognosen als quantitative Entscheidungsgrundlage für die Regionalplanung ableiten lässt. Insbesondere stochastische Modellierungen, die Methoden aus der Zeitreihenanalyse nutzen, erfreuen sich größer werdender Beliebtheit. Allerdings erfordern diese Ansätze möglichst lange Zeitreihen verschiedener demographischer Merkmale wie etwa der Sterblichkeit und Informationen über deren alters- und geschlechtsdifferenzierte Struktur, die auf regionaler Ebene häufig nicht verfügbar sind. Daher sind stochastische Modellierungen meist nur auf nationaler Ebene durchführbar. Darüber hinaus werden die für eine Bevölkerungsprognose erforderlichen Daten von den verschiedenen Statistischen Landesämtern häufig unterschiedlich aufbereitet. So weicht beispielsweise die Struktur von altersspezifischen Merkmalen zwischen den Ländern voneinander ab, wodurch die Bevölkerungsentwicklung gerade für länderübergreifende polyzentrische Metropolregionen schwer zu modellieren ist.

Der vorliegende Beitrag liefert eine stochastische Prognose der Bevölkerung der Metropolregion Rhein-Neckar gegliedert nach Altersjahren (bis 90+ Jahre) und Geschlecht von 2010 bis 2030. Die Region liegt im Schnittpunkt der drei Bundesländer Baden-Württemberg, Hessen und Rheinland-Pfalz und bildet einen gewachsenen Wirtschaftsraum. Dessen soziokulturelle Beziehungen reichen einerseits bis weit in die Vergangenheit zurück, in der der Raum als „Kurpfalz“ eine territoriale Einheit bildete. Andererseits haben sich seit den 1970er Jahren bundeslandübergreifende ökonomische Verflechtungen entwickelt, die mit der Initiative „Rhein-Neckar-Dreieck“ verbunden sind (*Lowack* 2007: 132). Die drei zentral gelegenen Oberzentren Mannheim, Heidelberg und Ludwigshafen prägen die polyzentrische Struktur der Region, die mit etwa 2,4 Millionen Einwohnern siebtgrößter Wirtschaftsraum Deutschlands ist. Das Gebiet erstreckt sich über insgesamt 15 Landkreise und kreisfreie Städ-

te. Metropolregionen gelten als „Motoren der wirtschaftlichen Entwicklung“ (*Bundesamt für Bauwesen, Raumordnung und Städtebau* 1995), von denen Impulse auf strukturschwächere Regionen ausstrahlen sollen. Sie stehen jedoch durch den zukünftigen Bevölkerungsrückgang in Konkurrenz, insbesondere um hoch qualifizierte Arbeitskräfte (*Klein* 2008: 44). Dieser „Kampf um kluge Köpfe“ (*Chambers et al.* 1998: 1) verdeutlicht die Bedeutung regionaler Bevölkerungsprognosen, die helfen, zukünftige Engpässe frühzeitig zu identifizieren und eine angemessene Reaktion zu ermöglichen, beispielsweise durch Aus- und Weiterbildungsprogramme (*Gans et al.* 2009: 118). Die Metropolregion Rhein-Neckar integriert die sich abzeichnenden Herausforderungen über einen „Lenkungsreis Demographie“, ein Netzwerk regionaler Entscheidungsträger, in die Regionalstrategie „Demographischer Wandel“ (*Metropolregion Rhein-Neckar* 2011: 26).

Die im Folgenden vorgestellte Bevölkerungsprognose nutzt das Paradigma funktionaler Daten für eine tief greifende Analyse der demographischen Entwicklung der Metropolregion Rhein-Neckar. Dieser Ansatz basiert auf der Idee, dass die sekundärstatistische Datenbasis der einzelnen zu Altersgruppen aggregierten demographischen Merkmale einem messbaren funktionalen Zusammenhang folgt. Ziel der Untersuchung ist es, diesen zu schätzen und für Prognosen zu nutzen. Auf diese Weise können die unterschiedlich strukturierten und gegliederten Datensätze der eingehenden demographischen Merkmale der statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz auf eine Untersuchungsregion geglättet und zu einer Gesamtregion aggregiert werden. Auf dem Paradigma funktionaler Daten basieren Zeitreihenmodelle (Kapitel 3), welche die Unsicherheit der demographischen Entwicklung durch Konfidenzintervalle genauer abbilden als die häufig verwendete Szenariotechnik. Kapitel 4 verdeutlicht die Vorteile des Ansatzes am Beispiel der Metropolregion Rhein-Neckar und Kapitel 5 schließt mit einem Ausblick.

2 Bevölkerungsprognosen

Die demographische Zukunft einer Region ist unsicher, denn es gibt sehr viele denkbare Verläufe. Die Literatur unterscheidet zwei methodische Ansätze, um mit dieser Herausforderung umzugehen. Die traditionelle Herangehensweise stellen deterministische Ansätze dar. Diese werden zwar häufig von der amtlichen Statistik genutzt, bilden die Entwicklung der Bevölkerung jedoch ohne genaue Informationen über die Unsicherheit über die zukünftige Entwicklung ab (*Lipps/Betz* 2004: 1). Eine Möglichkeit hierzu bieten mathematische Extrapolationsmethoden, die die Trends der Vergangenheit analysieren und über Wachstumsfunktionen fortschreiben. Die Verwendung von Wachstumsraten eignet sich dabei jedoch insbesondere für großräumige Vorausberechnungen wie beispielsweise die der Weltbevölkerung (*O'Neill et al.* 2001: 207) und weniger für eine Modellierung auf regionaler Ebene, deren Entwicklung durch eine größere Dynamik, insbesondere durch Wanderungsbewegungen, geprägt ist. Um diese Aspekte zu berücksichtigen, nutzen gängige deterministische Verfahren die demographische Grundgleichung als Grundlage (*Bähr et al.* 1992: 327):

$$B_t = B_{t-1} + G_{t-1,t} - S_{t-1,t} + M_{t-1,t}$$

mit B_t = Bevölkerung zum Zeitpunkt t ;
 B_{t-1} = Bevölkerung zum Zeitpunkt $t-1$;
 $G_{t-1,t}$ = Geburten zwischen $t-1$ und t ;
 $S_{t-1,t}$ = Sterbefälle zwischen $t-1$ und t sowie
 $M_{t-1,t}$ = Nettomigration zwischen $t-1$ und t .

Aus einer Ausgangsbevölkerung und Annahmen über die demographischen Determinanten (Fertilität, Mortalität und Nettomigration) berechnet sich ein möglicher Verlauf der zukünftigen Bevölkerung (*Statistisches Bundesamt* 2009: 9). Die demographische Grundgleichung unterscheidet verschiedene Zeitpunkte, betrachtet die Bevölkerung allerdings nur im Aggregat. Um die vielfältigen Veränderungen des demographischen Wandels sichtbar zu machen, muss die Grundgleichung nach Alter ($x = 0, 1, 2, \dots, k-1, k$) und Geschlecht (M=Männer und F=Frauen) der Bevölkerung differenziert werden. Die Geburten und Sterbefälle der einzelnen Altersklassen ergeben sich aus dem Produkt der zugrunde liegenden demographischen Ziffern mit der Bevölkerung des Vorjahres. Die demographische Entwicklung einer Region ist somit durch fünf Komponenten bestimmt: die Fertilitätsziffer, die männliche und weibliche Mortalitätsziffer sowie die geschlechtsdifferenzierte Nettomigration. Die Kohorten-Komponenten-Methode formalisiert die weibliche Bevölkerungsentwicklung (Index „F“) zu einem Matrixmodell (*Lee/Tuljapurkar* 1994: 1178):

$$B_{t+1}^F = \Omega_t^F B_t^F + M_{t+1,t}^F,$$

beziehungsweise (*Pflaumer* 1988: 136):

$$\begin{pmatrix} B_{0,t+1}^F \\ B_{1,t+1}^F \\ B_{2,t+1}^F \\ \vdots \\ \vdots \\ B_{k,t+1}^F \end{pmatrix} = \begin{pmatrix} 0 & \dots & v f_{15,t}^F \alpha_{*,t}^F & \dots & v f_{49,t}^F \alpha_{*,t}^F & 0 & \dots & 0 \\ \alpha_{0,t}^F & 0 & 0 & \dots & \dots & 0 & \dots & 0 \\ 0 & \alpha_{1,t}^F & 0 & \dots & \dots & 0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & & & & \vdots \\ \vdots & \vdots & & & 0 & \ddots & & \vdots \\ 0 & 0 & \dots & \dots & \dots & 0 & \alpha_{k-1,t}^F & 0 \end{pmatrix} \begin{pmatrix} B_{0,t}^F \\ B_{1,t}^F \\ B_{2,t}^F \\ \vdots \\ \vdots \\ B_{k,t}^F \end{pmatrix} + \begin{pmatrix} M_{0,t}^F \\ M_{1,t}^F \\ M_{2,t}^F \\ \vdots \\ \vdots \\ M_{k,t}^F \end{pmatrix}.$$

Die (weibliche) Bevölkerung eines Altersjahres x der Folgeperiode $t+1$ berechnet sich aus dem Summenprodukt einer Zeile der Matrix Ω (der Mortalitätsziffer $m_{x,t}$ bzw. Fertilitätsziffer $f_{x,t}$) mit dem Spaltenvektor der Bevölkerung zum Zeitpunkt t . Der Parameter v quantifiziert den Anteil der neugeborenen Mädchen und $\alpha_{x,t}$ die Überlebenswahrscheinlichkeit, definiert als $(1-m_{x,t})$. Dieses Produkt liefert einen Vektor für die natürliche Bevölkerungsentwicklung. Die Anzahl der Neugeborenen berechnet sich aus der Summe der Frauen im gebärfähigen Alter multipliziert mit den Fertilitätsziffern. Dieser Wert wird auf die beiden Geschlechter verteilt und anschließend

mit der Wahrscheinlichkeit, das erste halbe Lebensjahr zu überleben, multipliziert ($\alpha_{*,t}$). Aus der Addition des Vektors für die natürliche Bevölkerungsentwicklung mit dem Vektor der Nettomigration resultiert die Bevölkerung zum Zeitpunkt $t+1$.

Da die Kohorten-Komponenten-Methode die einzelnen Komponenten der demographischen Grundgleichung getrennt betrachtet und Kohorten (meist Altersjahre) unterscheidet, können Einflüsse, die nur auf bestimmte Altersbereiche einwirken, präzise abgebildet werden. Dieser Ansatz wurde bereits in den 1920er Jahren von *Whelpton* (1928) zur Vorausberechnung der Bevölkerung der USA von 1925 bis 1975 genutzt (*Bähr et al.* 1992: 500). Dieser Ansatz gilt heute als das Standardverfahren für Modellierungen der Bevölkerungsentwicklung und wird von der amtlichen Statistik genutzt (beispielsweise *Statistisches Bundesamt* 2009). Um die Unsicherheit der demographischen Zukunft zu berücksichtigen, werden durch verschiedene Annahmen über die demographischen Ziffern unterschiedliche Verläufe berechnet. Diese Szenarien (*Lutz et al.* 1998a: 140) umfassen meist eine optimistische, neutrale und pessimistische Entwicklung. Bei der Interpretation der Szenarien wird häufig der Fehler begangen, den Schwankungsbereich zwischen dem pessimistischen und optimistischen Verlauf als Maß für die Unsicherheit zu interpretieren (*Keyfitz* 1972: 353). Für regionale Prognosen erscheint die Kombination verschiedener Annahmen als ungeeignet, da insbesondere die zukünftigen Wanderungsbewegungen einen großen Einfluss auf die Bevölkerungsstruktur besitzen, aber mit sehr großer Unsicherheit behaftet sind.¹

Obwohl die deterministische Herangehensweise seit Jahren von den offiziellen Stellen praktiziert wird, ist sie nicht unproblematisch (*Lee* 1999; *Lipps/Betz* 2003; *Keilman et al.* 2002). Außer der Spannweite möglicher Entwicklungen trifft der deterministische Ansatz keine Aussagen über den Eintritt der einzelnen Szenarien. Stochastische Ansätze dagegen ermöglichen Prognosen, bei denen die Wahrscheinlichkeit angegeben wird, mit der das Ergebnis innerhalb eines bestimmten Schwankungsbereichs liegt und dies stellt beispielsweise für die Politikberatung eine wertvolle Zusatzinformation dar. So würden Regionalplaner gerne wissen, welches Szenario das wahrscheinlichste ist, um „böse Überraschungen“ zu vermeiden und konkretere Pläne zu formulieren. Darüber hinaus spricht die Annahme fixer demographischer Komponenten gegen die Verwendung von Szenarien. So wird die Anzahl der lebendgeborenen Kinder einer Frau häufig mit 1,4 festgelegt, wodurch die zukünftigen Werte vollkommen durch die vorangegangenen Perioden bestimmt sind. *Lipps* und *Betz* (2003: 4-5) zeigen an einem einfachen Rechenbeispiel, dass die Entwicklung einer als fix (perfekte Autokorrelation) angenommenen demographischen Komponente bei einem Zeithorizont von 20 Jahren deutlich von einer auf Zeitreihenmodellen basierenden Prognose trotz marginaler Unterschiede in den einzelnen Perioden abweicht. Darüber hinaus resultiert perfekte Autokorrelation zwischen den demographischen Variablen auch aus der Kombination der getroffenen Annahmen. Ein mögliches Szenario „hohes Bevölkerungswachstum“

¹ Darüber hinaus findet der interessierte Leser bei *Rogers* (1985) eine ausführliche Darstellung und Diskussion verschiedener Ansätze für regionale Bevölkerungsprognosen.

entsteht aus der Kombination hoher Fertilität und niedriger Sterblichkeit bzw. hoher Lebenserwartung. In diesem Szenario impliziert eine hohe Geburtenziffer eine hohe Lebenserwartung (*Keilman et al.* 2002: 410). Diese Autokorrelation zwischen den demographischen Variablen erzeugt Inkonsistenzen, da extreme Annahmen über eine Variable nicht zwangsläufig extreme Annahmen über eine andere Variable nach sich ziehen müssen. Eine Bevölkerungsprognose sollte daher zwei Bestandteile aufweisen (*Keilman et al.* 2002: 410):

1. eine Spannweite der möglichen Entwicklung und
2. eine Eintrittswahrscheinlichkeit für diese Spannweite.

Außerdem sollte eine Prognose vier Arten an Korrelation berücksichtigen können: über die Zeit, zwischen den verschiedenen Altersklassen, zwischen den Geschlechtern und zwischen den demographischen Komponenten.² Eine mögliche Korrelation zwischen den verschiedenen demographischen Ziffern spielt in entwickelten Ländern jedoch nur eine untergeordnete Rolle (*Keilman et al.* 2002: 412). Nur die stochastische Herangehensweise kann diese Anforderungen an eine Prognose erfüllen, denn die zukünftigen Werte liegen in einem Konfidenzintervall, das als Maß für die erwartete Genauigkeit dient. Dabei können drei Ansätze unterschieden werden (*Lipps/Betz* 2003: 5):

- Analyse historischer Prognosefehler
- Annahmen von Expertengruppen
- Zeitreihenmodelle

Die Analyse historischer Prognosefehler geht auf *Keyfitz* (1981) und *Stoto* (1983) zurück. Aus dem Vergleich früherer Prognosen mit der tatsächlichen Entwicklung können Prognosefehler abgeleitet werden, aus denen sich ein Prognoseintervall für die Zukunft bestimmen lässt (*Keilman et al.* 2002: 415). Dieser Ansatz liefert aber nur brauchbare Resultate, wenn die zukünftigen Schwankungen der demographischen Ziffern den bisherigen ähneln und Fehleinschätzungen auch für die Zukunft gelten (*Lipps/Betz* 2003: 5). Außerdem stehen häufig nur wenige Perioden zur Ableitung der Fehler und der Konstruktion der Prognoseintervalle zur Verfügung, weswegen die Güte des Ansatzes fragwürdig erscheint. Darüber hinaus basieren die frühere Prognosen auf Methoden, die häufig nicht mehr dem Stand der Forschung entsprechen und deshalb als ungenauer eingestuft werden müssen als Prognosen mit aktuellen Methoden. Auf regionaler Ebene ist dieser Ansatz aufgrund fehlender früherer Prognosen in der Regel nicht oder nur schwer umsetzbar.

Lutz et al. (1996, 1998a/b) nutzen die Annahmen von Expertengruppen zur Projektion der demographischen Raten und deren Unsicherheit (*Lutz et al.* 1996). Eine Expertengruppe einigt sich sowohl auf eine Punktschätzung als auch auf ein Konfi-

² Eine ausführliche Diskussion der verschiedenen Arten der Korrelation findet sich bei *Keilman et al.* (2002: 412-414).

denzintervall. Unter einer Verteilungsannahme für die demographischen Ziffern und Annahmen über deren Korrelation wird im nächsten Schritt durch eine Simulation eine Projektion der Bevölkerungsentwicklung mit einem Schwankungsbereich berechnet. Gegen die Verwendung dieser Herangehensweise führen *Keilman et al.* (2002: 415) an, dass selbst Experten nicht zuverlässig zwischen einem 95 %- und einem 99 %-Konfidenzintervall unterscheiden können. Außerdem ist das Risiko einer seriellen Korrelation hoch, die zu inkonsistenten Konfidenzintervallen führt. *Lee* (1999: 172) diskutiert, dass es keinen objektiven Weg gibt, um die Unsicherheit über zukünftige demographische Entwicklungen aus Expertenmeinungen abzuleiten. Darüber hinaus stehen geeignete Expertengruppen auf regionaler Ebene eventuell nicht zur Verfügung, so dass dieser Ansatz unter Umständen nicht umgesetzt werden kann.

Zeitreihenmodelle basieren auf der Annahme, dass die demographischen Entwicklungen der Vergangenheit durch ein statistisches Modell erklärt werden können und dieser Zusammenhang auch in der Zukunft gilt. Das wohl bekannteste Zeitreihenmodell ist das von *Lee* und *Carter* (1992) zur Modellierung der Mortalität. Dieser zweistufige Ansatz zur Abschätzung zukünftiger Trendentwicklungen der alters- und geschlechtsspezifischen Sterbeziffern passt im ersten Schritt das Modell an die Zeitreihen der Sterbeziffern an, um im zweiten Schritt deren Entwicklung in die Zukunft zu übertragen. Da der Ansatz leicht zu implementieren ist und sich als flexibel erwiesen hat, ist er vielfach adaptiert und weiterentwickelt worden (*Booth* 2006). *Hyndman* und *Booth* (2008: 324) geben einen Überblick über weitere Ansätze zur Modellierung der demographischen Raten. Zeitreihenmodelle eignen sich nur für kurz- und mittelfristige Prognosen (*Keilman et al.* 2002: 414), ein zu langer Zeithorizont resultiert in unrealistischen Prognosen und sehr weiten Konfidenzintervallen (*Sanderson* 1995: 274). Die wesentlichen Vorteile von Zeitreihenmodellen gegenüber den beiden stochastischen Alternativen für eine regionale stochastische Bevölkerungsprognose bestehen darin, dass die Unsicherheit konsistent abgebildet werden kann und weder frühere Prognosen noch Expertengruppen erfordern. Da auf regionaler Ebene viele Daten häufig nur für aggregierte Altersgruppen verfügbar sind, erscheinen das Paradigma funktionaler Daten und darauf aufbauende Modelle als hilfreich für eine regionale Bevölkerungsprognose.³

Denn dieser Ansatz ermöglicht eine effektive Disaggregation der Sekundärdaten. Die darauf aufbauenden Modelle von *Hyndman* und *Ullah* (2007) sowie *Hyndman et al.* (2011) weisen geringere Prognoseintervalle auf als alternative Ansätze (*Hyndman/Ullah* 2007: 4953; *Hyndman et al.* 2011: 25). Deshalb bilden diese Modelle die Grundlage für die stochastische Bevölkerungsprognose der Metropolregion Rhein-Neckar.

³ Der interessierte Leser findet eine ausführliche Diskussion der verschiedenen methodischen Ansätze bei *O'Neill et al.* (2001: 210-222).

3 Das Paradigma funktionaler Daten

3.1 Grundidee

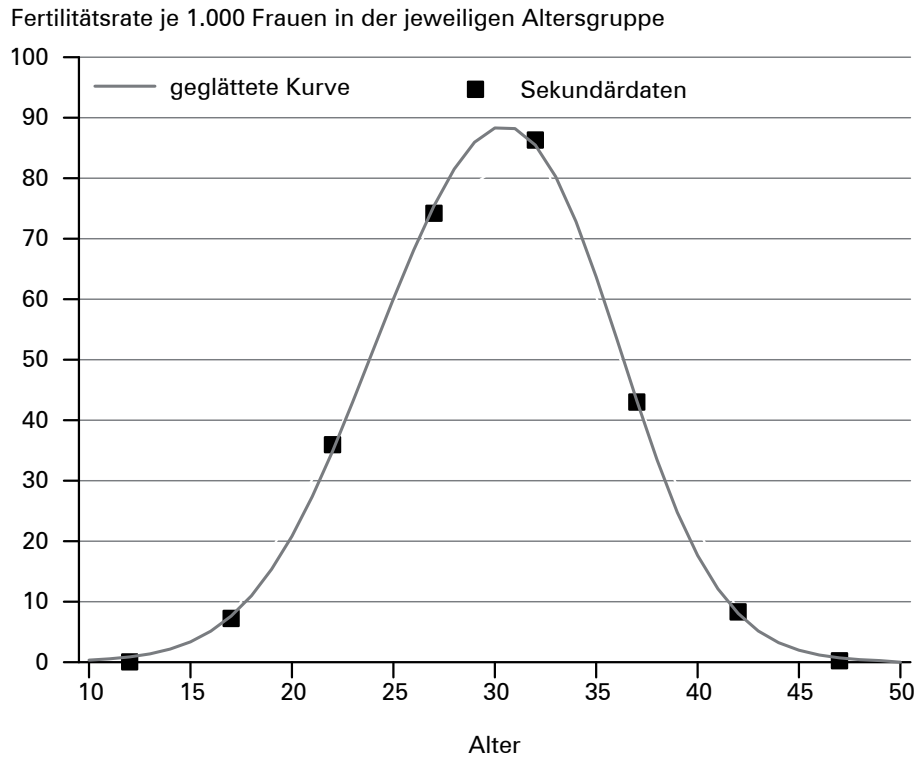
Eine tief greifende Analyse der Auswirkungen des demographischen Wandels auf die Bevölkerungsstruktur erfordert nach Altersjahren gegliederte demographische Merkmale. Bei einer nicht optimalen Datenlage erweist sich das Paradigma funktionaler Daten als überaus hilfreich, denn insbesondere auf regionaler Ebene sind Sekundärdaten häufig nur differenziert nach Altersgruppen verfügbar. Die Analyse funktionaler Daten („functional data analysis“, nachfolgend FDA) stellt einen Ansatz dar, der sich als Ansatz im Umgang mit Datenreihen begreift, bei dem eine Kurve einzelne Datenpunkte verbindet, eine zusammenhängende Reihe darstellen (*Ramsay/Silverman* 2001: 5822). Ziel der FDA ist es, diese Funktion zu schätzen und als Grundlage für Prognosen zu nutzen. Hierbei dient jede sekundärstatistische Beobachtung $Y_t(x)$ einer demographischen Komponente als Stützpunkt für eine glatte Funktion $s_t(x)$. Bei einem Horizont von N Zeitpunkten ergeben sich somit N Kurven. Jede Beobachtung unterliegt einem Messfehler $\varepsilon_t(x)$, woraus folgendes Modell resultiert (*Ramsay* 2008: 5):

$$Y_t(x) = s_t(x) + \varepsilon_t(x)$$

Um die für die FDA benötigten Informationen über die glatten Funktionen $s_t(x)$ zu erhalten, müssen die einzelnen demographischen Parameter mit ihren Informationen für Altersgruppen auf Altersjahre disaggregiert werden. Das statistische Glätten (smoothing) ist ein Ansatz, der versucht, wesentliche, aber nicht verfügbare Muster eines Datensatzes zu schätzen und dabei eventuelles Rauschen zu beseitigen. Nach *Wood* (1994: 27) liefern kubische Spline-Funktionen die besten Resultate für demographische Daten. Die Mediane der einzelnen Altersgruppen bilden Stützpunkte oder Knoten, von denen sich der erste am Anfang der ersten Altersgruppe (Alter=0) befindet, der nächste Knoten am Median der zweiten Gruppe und so weiter. Der letzte Knoten befindet sich am Ende der obersten Altersgruppe. Die Altersjahre, die zu einer Altersgruppe zusammengefasst sind, bilden einen Abschnitt. Das Glättungsverfahren vereinigt die Knoten durch unterschiedliche kubische Polynome für die einzelnen Abschnitte zu einer Kurve, die für alle Punkte stetig differenzierbar ist. Durch den Verlauf der Spline-Funktionen lässt sich für jedes Altersjahr ein Wert für die zu modellierende demographische Komponente ablesen. Abbildung 1 verdeutlicht das Vorgehen an einer beispielhaften Fertilitätsziffer: die Medianwerte der einzelnen Altersgruppen bilden die Knoten (Quadrate), zwischen denen kubische Splines verlaufen, aus denen sich die geglättete Kurve zusammensetzt.

Funktionale Daten, beziehungsweise die einzelnen glatten Funktionen für die N Zeitpunkte, unterliegen zwei Variationen: einem Amplitudenrauschen und einer Phasenverschiebung (*Ramsay* 2008: 3). Das Amplitudenrauschen stellt eine vertikale Variation der Funktionen über die Zeit dar, die „Höhe“ der Funktionswerte ist demnach nicht konstant. Die Phasenverschiebung zeigt die horizontale Variation an, dass zum Beispiel über N Zeitpunkte die altersspezifischen Ausprägungen der de-

Abb. 1: Glättung der Sekundärdaten am Beispiel der altersspezifischen Fertilitätsraten eines Jahres



Quelle: Eigener Entwurf.

mographischen Parameter einen bestimmten Trend aufweisen. So erhöhte sich in den letzten Jahrzehnten kontinuierlich das Alter der Frauen, für das das Maximum der altersspezifischen Geburtenziffer erreicht wurde. Diese Variationen der Kurven erschweren die Modellierung durch Zeitreihenmodelle, denn durch die Glättung der Daten auf Altersjahre wären Modelle mit sehr vielen Parametern nötig, um alle Entwicklungen im Alter und der Zeit adäquat darzustellen. Deshalb nutzt die FDA den sogenannten „time-warping“-Ansatz, um die charakteristischen Merkmale einer Funktion mit einer geringen Anzahl an Parametern abzubilden und deren Variation durch das Amplitudenrauschen und die Phasenverschiebung zu minimieren (Ramsay/Silverman 2001: 5823): Eine Linearkombination von Basisfunktionen φ_k ($k=1, \dots, K$) korrigiert die Verzerrungen:

$$\sum_{k=1}^K \beta_k \varphi_k(x).$$

Die Basisfunktionen enthalten Informationen über die alters- und zeitspezifischen Entwicklungen der glatten Funktion. Ein Zeitreihenmodell wird umso flexi-

bler, je mehr Basisfunktionen verwendet werden. Die große Herausforderung besteht darin, die Funktion einfach und die Anzahl der verwendeten Parameter gering zu halten. Denn je mehr Basisfunktionen in die Berechnung der glatten Funktion eingehen, umso sensibler reagiert die Funktion auf Messfehler. Dieser Ansatz ist weit verbreitet, da die Basisfunktionen durch eine Hauptkomponentenanalyse einfach geschätzt werden können, welche die Entwicklung aller wesentlichen Merkmale der zugrunde liegenden Kurven im Zeitverlauf beschreibt. Diese auch als Registrierung bezeichnete Ausrichtung der Funktionen über den Zeithorizont erleichtert weiterführende Auswertungen und verbessert die Vergleichbarkeit zwischen verschiedenen Zeitpunkten. Die verbleibende Variation der Kurven geht nur noch auf reines Amplitudenrauschen zurück.

3.2 Modelle

Hyndman und Ullah (2007) nutzen das Paradigma funktionaler Daten zur Entwicklung eines Modells, das verschiedene demographische Parameter zuverlässig abbilden kann. So bezeichnet $Y_t(x)$ den Wert einer demographischen Rate für das Alter x zum Zeitpunkt t . Auf Basis der Grundidee der FDA liegt dem Modell eine glatte Funktion $s_t(x)$ zugrunde, die mit einem Fehler zu einem diskreten Zeitpunkt in Abhängigkeit vom Alter beobachtet wird. Für die (sekundärstatistischen) Datenpunkte $\{x_i, y_t(x_i)\}$ mit $t = 1, \dots, n$ und $i = 1, \dots, \text{„höchste Altersklasse“}$, gilt:

$$Y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_t(x).$$

Die Fehlerterme $\varepsilon_{t,x}$ sind unabhängig und identisch verteilt und durch den Term σ_t in Abhängigkeit vom Alter gewichtet. Dies ist sinnvoll, denn bei kleinen Funktionswerten in bestimmten Altersbereichen, wie beispielsweise bei der Fertilitätsziffer für Frauen unter 15 und über 45 Jahren, sind auch die Messfehler entsprechend klein. Ohne Gewichtung könnte der Eindruck entstehen, dass das Modell in diesen Abschnitten die Daten besser abbildet als in anderen Altersbereichen. Die Modellierung zielt auf die Prognose von $Y_t(x)$ für den Horizont von h Zeitpunkten ($t = n+1, \dots, n+h$) und alle definierten Altersklassen. Dieser Ansatz basiert auf dem Paradigma funktionaler Daten und versucht das Amplitudenrauschen und die Phasenverschiebung als Quelle für Messfehler zu identifizieren und zu erfassen:

1. Für eine Bevölkerungsprognose sind Daten wünschenswert, die nach Altersjahren gegliedert sind, denn nur so können zum Beispiel die vielfältigen Wirkungen des demographischen Wandels auf die Altersstruktur der Bevölkerung sichtbar gemacht werden. Häufig sind sekundärstatistische Daten allerdings zu Altersgruppen aggregiert. In diesem Fall wird mit Hilfe nicht-parametrischer Smoothing-Methoden für jeden vorliegenden Zeitpunkt die Altersstruktur geglättet, um die Y_t zugrunde liegende Funktion schätzen zu können. Für jeden Zeitpunkt wird eine separate Funktion geschätzt.
2. Die Modellierung aller wesentlichen charakteristischen Eigenschaften eines demographischen Merkmals erfordert eine Vielzahl an Parametern. Die hierfür benötigten langen Zeitreihendaten sind insbesondere auf regionaler Ebe-

ne, unter anderem aufgrund von Gebietsreformen, selten verfügbar. Stattdessen zerlegen Basisfunktionen die in Schritt eins geschätzten Funktionen in Einzelbausteine, die alle wesentlichen Merkmale enthalten, dabei aber mit einer geringen Anzahl an Parametern auskommen:

$$s_t(x) = \mu(x) + \sum_{k=1}^K \beta_{t,k} \varphi_k(x) + e_t(x),$$

wobei $\mu(x)$ den Mittelwert von $s_t(x)$ über alle betrachteten Jahre darstellt, $\{\varphi_k(x)\}$ ist ein Set an Basisfunktionen und $e_t(x)$ ein normalverteilter Fehlerterm mit Mittelwert 0 und Varianz $\text{var}(x)$. Die Basisfunktionen sind das Ergebnis einer Hauptkomponentenanalyse. Der Mittelwert bildet die wesentliche Struktur des modellierten demographischen Merkmals ab, während die im Alter und der Zeit vom Mittelwert abweichenden Entwicklungen durch das Amplitudenrauschen und die Phasenverschiebung in den Basisfunktionen enthalten sind.

3. Für jeden Koeffizienten $\beta_{t,k}$ mit $k = 1, \dots, K$ wird ein univariates Zeitreihenmodell geschätzt. *Hyndman* und *Booth* (2008: 327) zeigen, dass die Methode unempfindlich auf die Wahl von K reagiert, wenn ein hinreichend hoher Wert zugrunde liegt. Aus diesem Grund ist es sinnvoll, möglichst viele Basisfunktionen zu bilden. Allerdings steigt dadurch die Dauer der Berechnung der Modelle und die Basisfunktionen mit einem geringen Erklärungsgehalt sind unter Umständen nur schwer oder gar nicht interpretierbar.
4. Die Variation durch das Amplitudenrauschen und die Phasenverschiebung unterliegt zeitlichen Schwankungen. Der Parameter $\beta_{t,k}$ bildet diese Entwicklungen ab und wird für einen Zeithorizont von h Zeitpunkten prognostiziert:

$$\hat{y}_{n,h}(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{n,k,h} \hat{\varphi}_k(x),$$

wobei $\hat{\beta}_{n,k,h}$ den Schätzer zum Zeitpunkt $n+h$ für $\beta_{n+h,k}$ kennzeichnet.

5. Die geschätzten und prognostizierten Werte bilden die Grundlage für die Berechnung der zukünftigen Werte von $s_t(x)$ aus Schritt zwei, denn aus den Punktschätzern, multipliziert mit den geschätzten Basisfunktionen, berechnen sich die Prognosen für alle Altersgruppen.
6. Aus den geschätzten und prognostizierten Varianzen der Fehlerterme (Schritt zwei) berechnen sich die Prognoseintervalle. Diese Konfidenzintervalle stellen eine wertvolle Zusatzinformation gegenüber deterministischen Ansätzen dar, denn anstatt den Schwankungsbereich möglicher Entwicklungen einfach durch eine Variation von Szenarien mit unterschiedlichen Annahmen abzubilden, liegen diesem Ansatz statistische Methoden und Verteilungen zugrunde.

Der Vorteil dieses Ansatzes ist, dass er glatte Funktionen erlaubt, robust gegenüber Strukturbrüchen wie Krisen, Kriegen oder Katastrophen ist und einen Modellrahmen liefert, der Restriktionen zulässt,⁴ aber zeitgleich für verschiedene (demo-

⁴ Der Ansatz erlaubt qualitative Restriktionen, so kann beispielsweise die Annahme, dass die Wahrscheinlichkeit zu sterben ab einem gewissen Alter monoton ansteigt, integriert werden. Die im Rahmen dieser Arbeit verwendeten Restriktionen finden sich in Abschnitt 4.2.

graphische) Modellierungen nutzbar ist. So prognostizieren beispielsweise *Erbas et al.* (2007) das Risiko von Brustkrebserkrankungen australischer Frauen. Das Modell ist eine Verallgemeinerung des wegweisenden Lee-Carter-Modells (*Lee/Carter* 1992) zur Schätzung und Prognose von Mortalitätsziffern, das sich in vielen Erweiterungen auch auf andere demographische Merkmale anwenden lässt, zum Beispiel auf Fertilitätsdaten (*Lee* 1993). Die Grundidee des Lee-Carter-Modells findet sich auch in den FDA-Modellen wieder, allerdings unterscheidet sie sich in einigen Punkten, so setzten Lee und Carter unter anderem keine glatten Funktionen voraus (*Hyndman/Ullah* 2007: 326).

Der Ansatz von *Hyndman* und *Ullah* (2007) schätzt die Subpopulationen des modellierten demographischen Merkmales (häufig untergliedert in männlich und weiblich) unabhängig voneinander. Die prognostizierte Entwicklung beider Ziffern kann in diesem Fall bei einem langen Zeithorizont divergieren. Dieser Effekt lässt sich inhaltlich aber in der Regel nicht oder nur schwer erklären oder begründen. Für die Mortalitätsziffern beobachtet *Wilson* (2001: 167) eine globale Konvergenz sowohl zwischen verschiedenen Ländern als auch zwischen den Geschlechtern. Ansätze, die diesen Aspekt aufgreifen und explizit modellieren, werden in der Literatur als „coherent models“ (*Li/Lee* 2005: 581) bezeichnet. Prognosen, die auf dieser Denkweise basieren, sollen eine gewisse strukturelle Beziehung aufweisen. Für Mortalitätsziffern wurde dies durch Erweiterungen des klassischen Lee-Carter-Ansatzes methodisch beispielsweise von *Lee* und *Nault* (1993), *Lee* (2000) sowie *Li* und *Lee* (2005) umgesetzt. *Hyndman et al.* (2011) nutzten das Paradigma funktionaler Daten und erweitern den Ansatz von *Hyndman* und *Ullah* (2007), so dass eine divergierende Entwicklung als Folge von ungenauen Schätzungen vermieden werden kann. Da im Rahmen der stochastischen Bevölkerungsprognose eine Differenzierung nur nach Geschlechtern erfolgt, beschränken sich die nachfolgenden Ausführungen auf den Fall für zwei Subpopulationen. Der Ansatz erweist sich darüber hinaus als sehr flexibel und kann auf eine größere Zahl an Subpopulationen erweitert werden.

Den sekundärstatistischen Werten $Y_{t,F}(x)$ einer demographischen Komponente liegt eine glatte Funktion $s_{t,F}(x)$ zugrunde, die mit einem Fehler beobachtet wird. Bei zwei Subpopulationen gilt für die Frauen (Subskript F):

$$Y_{t,F}(x) = \log(s_{t,F}(x_i)) + \sigma_{t,F}(x)\varepsilon_t(x_i),$$

wobei x_i den Median der Altersgruppe i ($i = 1, \dots, p$) bezeichnet, $\varepsilon_t(x_i)$ ist eine identisch, unabhängig und gleichverteilte Zufallsvariable und $\sigma_{t,F}$ ein Term, durch den der Messfehler im Alter variiert. Bis auf die logarithmische Formulierung entspricht dieser Ansatz dem erläuterten Ansatz von *Hyndman* und *Ullah* (2007). Die glatte Funktion wird sowohl für Männer als auch Frauen beobachtet. Die Modellierung einer kohärenten Entwicklung erfordert die folgenden zwei Terme: das geometri-

sche Mittel (P_t für Produkt-Modell) und die Quadratwurzel aus dem Verhältnis (R_t für Verhältnis-Modell) jeweils für die geglätteten Ziffern einer demographischen Komponente $s_{t,F/M}(x)$:

$$P_t(x) = \sqrt{s_{t,M}(x) \cdot s_{t,F}(x)}$$

$$R_t(x) = \sqrt{s_{t,M}(x)/s_{t,F}(x)}.$$

Das Produkt-Modell beschreibt den altersspezifischen Verlauf eines demographischen Merkmals, während das Verhältnis-Modell die geschlechtsspezifische Variation anpasst. Bei Werten nahe eins von R_t erfolgen kaum Korrekturen, während von eins abweichende Werte geschlechtsspezifische Abweichungen in einem bestimmten Altersbereich bedeuten. Der Vorteil dieses Ansatzes ist, dass durch die logarithmische Formulierung aus dem Produkt in P_t eine Summe beziehungsweise aus dem Verhältnis in R_t eine Differenz wird. *Tukey* (1977) hat gezeigt, dass beide annähernd unkorreliert sind.

Auch das Produkt-Modell und das Verhältnis-Modell unterliegen Variationen durch das Amplitudenrauschen und die Phasenverschiebung. Um eine divergierende Entwicklung der demographischen Rate zu verhindern und gleichzeitig deren alters- und geschlechtsspezifische Veränderung abzubilden, werden P_t und R_t durch das Basisfunktionen-Modell für funktionale Daten von *Hyndman* und *Ullah* (2007) dargestellt:

$$\log[P_t(x)] = \mu_P(x) + \sum_{k=1}^K \beta_{t,k} \varphi_k(x) + e_t(x)$$

$$\log[R_t(x)] = \mu_R(x) + \sum_{l=1}^L \gamma_{t,l} \psi_l(x) + w_t(x).$$

Die Basisfunktionen $\varphi_k(x)$ und $\psi_l(x)$ beinhalten die vom Mittelwert (μ_P und μ_R) abweichenden Merkmale von P_t beziehungsweise R_t . Die Veränderung der Basisfunktionen über die Zeit bilden die Koeffizienten $\beta_{t,k}$ und $\gamma_{t,l}$ ab. Die Modellfehler messen die Residuen $e_t(x)$ und $w_t(x)$.

Um sicherzustellen, dass die Schätzer für die Prognose kohärent sind, also nicht divergieren, gilt für die zu verwendenden Zeitreihenmodelle, dass die Koeffizienten der Basisfunktionen $\{\beta_{t,k}\}$ und $\{\gamma_{t,l}\}$ stationäre Prozesse sind. Die prognostizierten Schätzer werden mit den Basisfunktionen multipliziert. Das Ergebnis sind Prognosen bis zum Zeitpunkt $n+h$ für zukünftige $P_t(x)$ und $R_t(x)$ aus denen sich die zukünftigen geschlechtsspezifischen Werte für die zu modellierende demographische Komponente ergeben:

$$Y_{n,h,M}(x) = P_{n,h}(x) \cdot R_{n,h}(x)$$

$$Y_{n,h,F}(x) = P_{n,h}(x) / R_{n,h}(x).$$

Die Prognose von $\beta_{n,k,h}$ und $\gamma_{n,l,h}$ für den Zeitpunkt $n+h$ für die einzelnen Geschlechter (Subskript j kennzeichnet „männlich“ oder „weiblich“) ist bestimmt durch:⁵

$$\log[y_{n,h,j}(x)] = \hat{\mu}_j(x) + \sum^K \hat{\beta}_{n,k,h} \varphi_k(x) + \sum^L \hat{\gamma}_{n,l,h} \psi_{l,j}(x).$$

Die vorgestellten Modelle auf Basis des Paradigmas funktionaler Daten eignen sich für kurz- und mittelfristige Prognosen von ca. 20 Jahren (*Hyndman/Booth* 2008: 339). Bei einem längeren Zeithorizont sinkt die Genauigkeit und die Breite der Konfidenzintervalle steigt. Die Prognosen der demographischen Merkmale basieren auf Informationen über die Entwicklung in der Vergangenheit, weswegen sie zukünftige Trendumkehrung nicht abbilden können. Beides trifft allerdings grundsätzlich auf die Verwendung von Zeitreihenmodellen zu und hat keinen Einfluss auf die Modellauswahl für die Prognose der Bevölkerung. Im direkten Vergleich zu alternativen Spezifikationen zeigen beide Ansätze eine größere Genauigkeit (vgl. *Hyndman/Ullah* 2007: 4953 und *Hyndman et al.* 2011: 17). Die Modellierung der geschlechtsdifferenzierten demographischen Merkmale (Nettomigration und Mortalität) erfolgt durch den Ansatz von *Hyndman et al.* (2011), während für die Fertilitätsziffer das Modell von *Hyndman und Ullah* (2007) unterstellt wird. Beide Ansätze eignen sich insbesondere für regionale Bevölkerungsprognosen, denn durch die Verwendung von Basisfunktionen sind die Ergebnisse sehr robust gegenüber Ausreißern und Strukturbrüchen. Dies ist ein Vorteil, da auf regionaler Ebene beispielsweise durch Gebietsreformen lange Zeitreihen häufig nicht verfügbar sind. Außerdem können gesellschaftliche Entwicklungen, wie die mögliche Trendumkehr von der Suburbanisierung hin zur Reurbanisierung, auf entsprechend niedriger Aggregationsebene durch die Basisfunktionen abgebildet werden. Darüber hinaus bieten die Statistischen Landesämter viele Merkmale nur gegliedert nach Altersgruppen an. Durch die Ideen des Paradigmas funktionaler Daten können diese Informationen auf Altersjahre disaggregiert werden. Die Basisfunktionen ermöglichen darüber hinaus auch die Modellierung von demographischen Trends und Strukturbrüchen, die sich in Form einer Phasenverschiebung oder eines Amplitudenrauschens in der Zeitreihe bemerkbar machen. Ob sie inhaltlich interpretiert werden können oder lediglich zur Korrektur statistischer Effekte dienen, hängt von ihrem Erklärungsgehalt ab.

4 Die Bevölkerungsentwicklung der Metropolregion Rhein-Neckar

4.1 Algorithmus

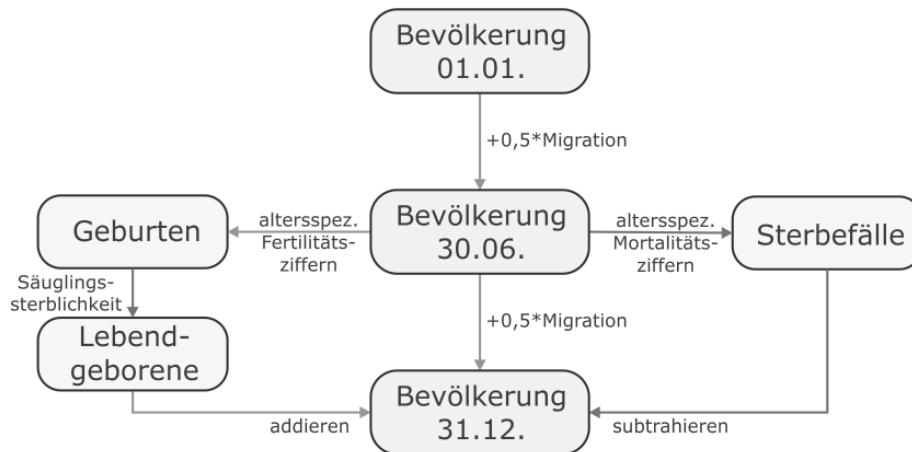
Fünf demographische Komponenten beeinflussen die Entwicklung der Bevölkerung: die Fertilitätsziffer, die geschlechtsspezifischen Mortalitätsziffern sowie die männ-

⁵ Das genaue Vorgehen zur Prognose der einzelnen Koeffizienten findet sich bei *Hyndman et al.* (2011: 6-7).

liche und weibliche Nettomigration. Die Kohorten-Komponenten-Methode formalisiert diesen Zusammenhang zu einem Gleichungssystem, allerdings ohne dabei genau die Vorgänge innerhalb eines Jahres zu spezifizieren. Die vorliegende Bevölkerungsprognose basiert auf dem Algorithmus von *Hyndman* und *Booth* (2008: 340), der nachfolgend schematisch dargestellt ist (Abb. 2). Ausgangspunkt ist die Bevölkerung zum 31.12. des Vorjahres. Die Wanderungsbewegungen ereignen sich nicht vollständig zu einem fixen Zeitpunkt, sondern über das Jahr verteilt. Während dieser Zeit können Zugewanderte theoretisch Kinder bekommen oder sterben. Der Algorithmus berücksichtigt diesen Aspekt, indem die Anzahl der simulierten altersspezifischen Wanderungsbilanz zur Hälfte auf den 01.01. und den 31.12. eines Jahres verteilt wird. Die Bevölkerung zum 30.06. eines Jahres entspricht der Bevölkerung zum 31.12. des Vorjahres zuzüglich der ersten Hälfte der Nettomigration. Die Sterbefälle berechnen sich aus der Bevölkerung zum 30.6. multipliziert mit den Mortalitätsziffern. Analog ergibt sich die Anzahl der Lebendgeborenen, korrigiert um die Säuglingssterblichkeit. Die Neugeborenen verteilen sich auf Basis des gleitenden Durchschnitts der letzten fünf Jahre auf die Geschlechter. Die weibliche Bevölkerung zum Zeitpunkt t+1 berechnet sich durch:

$$B_{t+1}^F = \Omega_t^F [B_t^F + 0,5 \cdot M_t] + 0,5 \cdot M_t.$$

Abb. 2: Algorithmus für die Kohorten-Komponenten-Methode



Quelle: Eigener Entwurf in Anlehnung an *Hyndman* und *Booth* (2008).

Die Bevölkerung zum Jahresende (31.12.) bestimmt sich aus der Summe der Bevölkerung zum 30.06. eines Jahres, den Sterbefällen, den bereinigten Lebendgeborenen und der zweiten Hälfte der Nettomigration. Diese Bevölkerung ist dann der Ausgangspunkt für die Berechnungen des Folgejahres, für das der Algorithmus von vorne durchläuft. Anders als deterministische Modelle, die pro Szenario jeweils eine einzelne Bevölkerungsentwicklung vorausberechnen, erfordern stochastische

Ansätze wegen der Unsicherheit über die zukünftige Entwicklung aufwändigere Simulationen, da die prognostizierten Werte jeweils um einen Mittelwert streuen. Auf Basis der unterstellten Verteilungen für die einzelnen demographischen Ziffern wird eine vorgegebene Anzahl möglicher Entwicklungspfade der zukünftigen Bevölkerung simuliert und in einer Datenbank gespeichert. Das Konfidenzintervall, in dem die zukünftige Bevölkerung liegt, reicht dabei vom $(100-\alpha)/2$ -ten Perzentil bis zum $(100+\alpha)/2$ -ten Perzentil (Keilman *et al.* 2002: 416).

Die Grundlage für die Prognose der Bevölkerung bilden getrennte Simulationen der Modelle aus Abschnitt 3.2 für die Fertilitätsziffer, die männliche und weibliche Mortalitätsrate sowie die geschlechtsspezifische Nettomigration jeweils für die Jahre 2010 bis 2030. Die Fertilitäts- und Mortalitätsziffern werden auf die Bevölkerung zum 30.06. bezogen (vgl. Abb. 2). Die Geburten und Sterbefälle folgen per Annahme der Poisson-Verteilung und berechnen sich durch zufälliges Ziehen aus der Verteilung und Aufsummierung dieser Werte über den relevanten Altersbereich. Hyndman und Booth (2008: 328) geben eine detaillierte Übersicht über die Generierung der jeweiligen Entwicklungspfade und deren Integration in die Bevölkerungsprognose.

Alle nachfolgenden Berechnungen wurden im Statistikprogramm „R“ (Version 2.12.0) durchgeführt. Diese freie Statistiksoftware ist auf allen gängigen Plattformen verfügbar und enthält im Zusatzpaket „demography“ auch die beschriebenen Ansätze für funktionale Daten sowie den vorgestellten Algorithmus für die Kohorten-Komponenten-Methode. Die Prognosen erfordern das Paket „forecast“. Insgesamt wurden 1000 Verläufe der Bevölkerungsentwicklung simuliert und anschließend zur Berechnung der Prognoseintervalle aus „R“ in einer Excel-Datenbank exportiert.

4.2 Daten: Quellen und Aufbereitung

Die für die Bevölkerungsprognose verwendeten Daten stammen von den Statistischen Landesämtern Baden-Württemberg, Hessen und Rheinland-Pfalz. Alle aggregierten Werte für die Metropolregion entstehen aus der Summe der Einzelwerte der 15 Landkreise beziehungsweise kreisfreien Städte. Die Berechnung der Fertilitätsziffer erfordert die Anzahl der Lebendgeborenen nach Altersgruppen der Mutter {10-14, 15-19, ..., 40-44, 45-49}. Da die Daten nicht zwischen Geschlechtern der Lebendgeborenen unterscheiden, kommen per Annahme auf 100 geborene Mädchen 105 geborene Jungen. Die Mortalitätsziffer berechnet sich aus der Anzahl der geschlechtsspezifischen Sterbefälle je Altersgruppe {<1, 1-4, 5-9, ..., 70-74, 75 und älter} zu der Anzahl der Männer oder Frauen in der entsprechenden Kohorte. Beide Merkmale liegen jeweils für die Jahre 1981 bis 2009 vor. Für die Berechnung der Raten werden diese Merkmale auf die Bevölkerungsdaten des entsprechenden Zeitraumes, gegliedert nach Altersgruppen {<3, 3-5, 6-9, 10-14, 15-17, 18-19, 20-24, 25-29, ..., 70-74, 75 und älter} und Geschlecht, bezogen. Als problematisch erwies sich die Berechnung der Mortalitätsziffern für Jungen und Mädchen unter 10 Jahren, da die Altersgruppen der Sterbefälle und Risikobevölkerung voneinander abweichen. Aus diesem Grund wurden die Bevölkerungsdaten einer Gruppe gleichmäßig auf die einzelnen Altersjahre verteilt und auf die Altersgruppen der Sterbefälle aggregiert. Die Daten zur Nettomigration für die Jahre 2002 bis 2009 sind nach Geschlecht

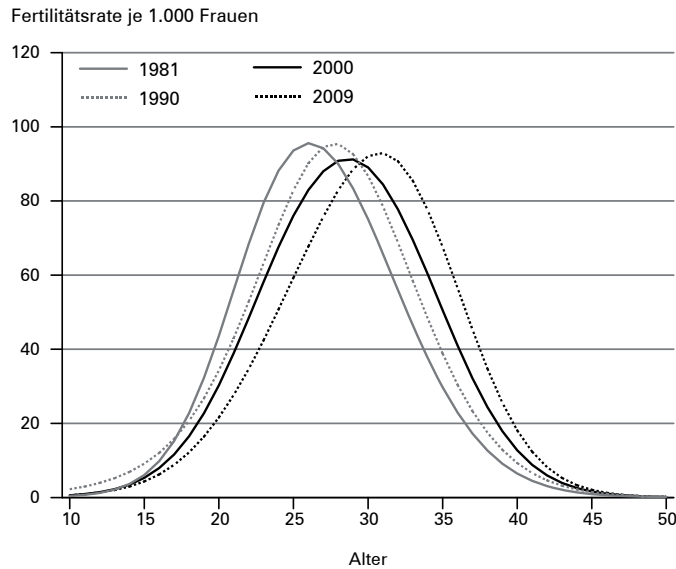
differenziert und in die Altersgruppen {<18, 18-24, 25-29, 30-49, 50-64, 65 und älter} unterteilt. Eigene Auswertungen auf Basis der vorgestellten Residualmethode ergänzen diese Reihen um die Jahre 1995 bis 2001.

Diese Datensätze erfüllen allerdings nicht die Anforderungen an die vorgestellten Modelle. Eine Analyse funktionaler Daten erfordert glatte (Alters-)Funktionen, die den Sekundärdaten zugrunde liegen. Für jede Altersgruppe der einzelnen demographischen Ziffern wird deshalb der Median gebildet und als Knoten genutzt, zwischen denen die geglätteten demographischen Ziffern verlaufen. Die Mortalitätsziffern sollen ab einem Schwellenwert (beispielsweise Altersjahre größer 50) monoton steigen (Hyndman/Ullah 2007: 4945), denn je älter ein Mensch ist, desto höher ist die Wahrscheinlichkeit, dass man stirbt. Diese qualitative Restriktion kann mit gewichteten „penalized regression splines“ nach Wood (1994) in die Glättung der Daten integriert werden. Dieses Vorgehen reduziert Verzerrungen der geschätzten Kurven in den oberen Alterskohorten. Aus demselben Grund werden die Datensätze der Fertilitätsziffern als konkav angenommen und mit gewichteten „B regression splines“ nach der Methode von He und Ng (1999) aufbereitet. Die Nettomigration ist kein Relativwert und kann sowohl positiv als auch negativ sein. Als Glättungsverfahren wird eine gewichtete „locally quadratic regression“ durchgeführt (Moore et al. 1997). Diese drei Glättungsverfahren sind im „demography“-Paket für „R“ enthalten.

Abbildung 3 zeigt die zeitliche Entwicklung der altersspezifischen Fertilitätsziffer für den Zeitraum zwischen 1981 und 2009 anhand der Kurven für 1981, 1990, 2000 und 2009. Die Verlagerung des Maximums der altersspezifischen Raten spiegelt die Zunahme des mittleren Alters der Mütter bei der Geburt von 26,9 Jahre (1981) auf 29,6 Jahre (2009) wider. Diese Entwicklung zeigt den Vorteil einer auf dem Paradigma funktionaler Daten basierender Modellierung auf, denn die empirischen Kurven für einen bestimmten demographischen Parameter weichen über den Horizont von N Zeitpunkten vom Mittelwert dieses Parameters ab. Die Entwicklung entspricht einer Phasenverschiebung, die bei der Modellierung der Zeitreihe durch die Basisfunktionen des FDA-Modells berücksichtigt wird.

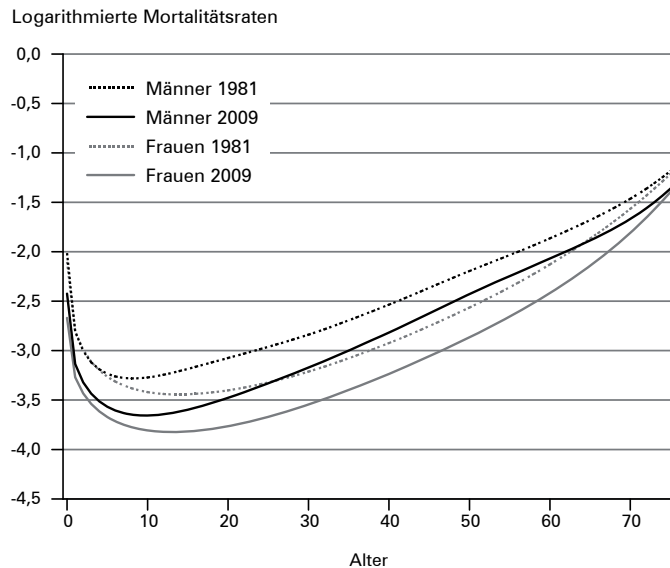
Der in vielen entwickelten Ländern und Regionen beobachtete Trend einer stetig ansteigenden Lebenserwartung findet sich auch in den Mortalitätsziffern der Männer und Frauen in der Metropolregion Rhein-Neckar (Abb. 4). Dabei haben Frauen über alle Altersjahre eine höhere Lebenserwartung bzw. geringere Mortalitätsziffern als Männer. Zwischen 1981 und 2009 sinken für beide Geschlechter die Ziffern gleichermaßen. Der wesentliche Unterschied findet sich in den männlichen Mortalitätsziffern zwischen 35 und 65 Jahren, die annähernd linear verlaufen, während die Kurven der Frauen eher einen konkaven Verlauf aufweisen. Die Mortalitätsziffern zwischen den Geschlechtern differieren in diesem Intervall am stärksten. In den höheren Altersklassen nähern sich die geschlechtsspezifischen Kurven gegenseitig an. Für alle Menschen über 75 Jahre konnten die (logarithmierten) Ziffern aufgrund der Datenlage nicht geglättet werden und mussten deshalb linear interpoliert werden. Die Zielwerte für die Altersgruppe „90+ Jahre“ der einzelnen Jahre stammen aus den Sterbetafeln des Statistischen Bundesamtes (*Statistisches Bundesamt*

Abb. 3: Geglättete altersspezifische Fertilitätsziffer je 1000 Frauen von 1981, 1990, 2000 und 2009



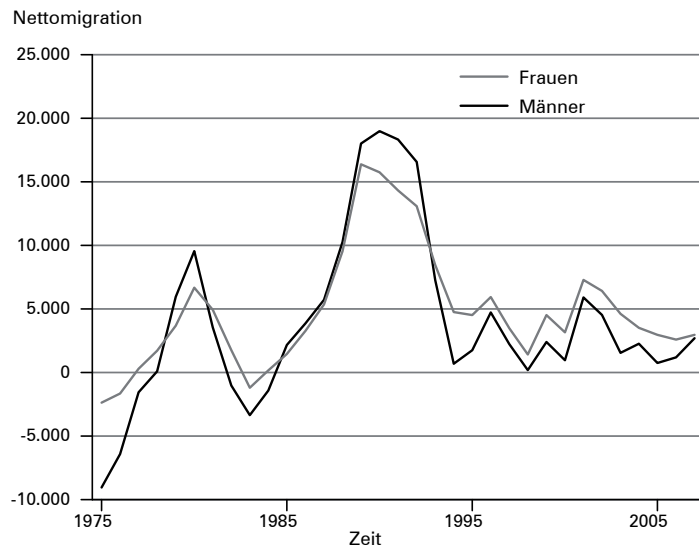
Quelle: Eigene Berechnung der funktionalen Daten unter Verwendung von „B regression splines“ nach der Methode von *He* und *Ng* (1999) und auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 4: Geglättete und logarithmierte Mortalitätsziffern für Männer und Frauen von 1981 und 2009



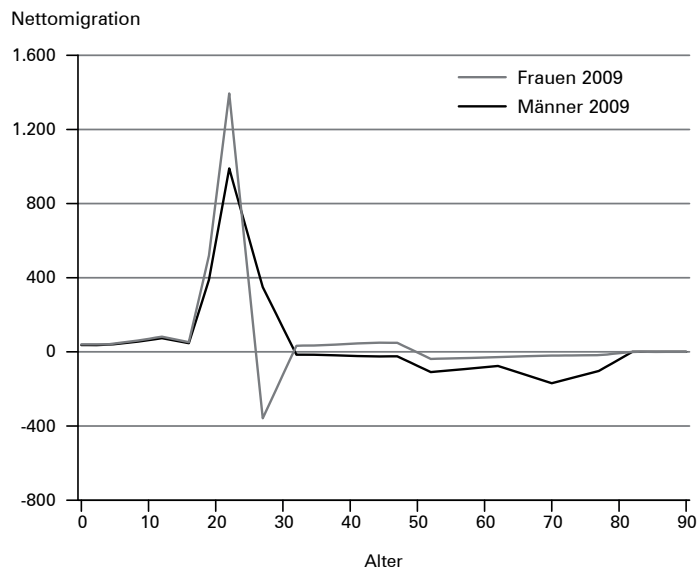
Quelle: Eigene Berechnung der funktionalen Daten unter Verwendung von gewichteten „penalized regression splines“ nach *Wood* (1994) und auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 5: Entwicklung der Nettomigration (von 1975 bis 2009)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 6: Geschlechtsdifferenzierte Nettomigration nach Altersjahren von 2009



Quelle: Eigene Berechnung der funktionalen Daten unter Verwendung von gewichteten „locally quadratic regression“ nach Moore *et al.* (1997) und auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

2010: 59). Die Bevölkerung im Alter ab 90 Jahre fällt in die oberste Altersklasse „90 und älter“.

Die Nettomigration unterliegt vielfältigen Einflüssen, die sowohl in der jeweiligen Herkunfts- als auch in der Zielregion wirksam sind und wodurch deutlich stärkere zeitliche Schwankungen als bei den Fertilitäts- und Mortalitätsraten entstehen (Abb. 5). Nachdem die Zu- und Fortzüge im Zusammenhang mit der deutschen Wiedervereinigung Anfang der 1990er Jahre abebbten, stabilisierten sich die Reihen merklich, und seit etwa Mitte der 1990er Jahre sind die Wanderungsüberschüsse in der Metropolregion von Frauen höher als die der Männer (Abb. 5). Die Daten zeigen, dass insbesondere junge Männer und Frauen im Alter zwischen 18 und 25 Jahre in die Metropolregion wandern (Abb. 6), was die Bedeutung als Ausbildungs- und Hochschulstandort verdeutlicht. Leichte Wanderungsverluste gibt es lediglich in den oberen Altersgruppen.

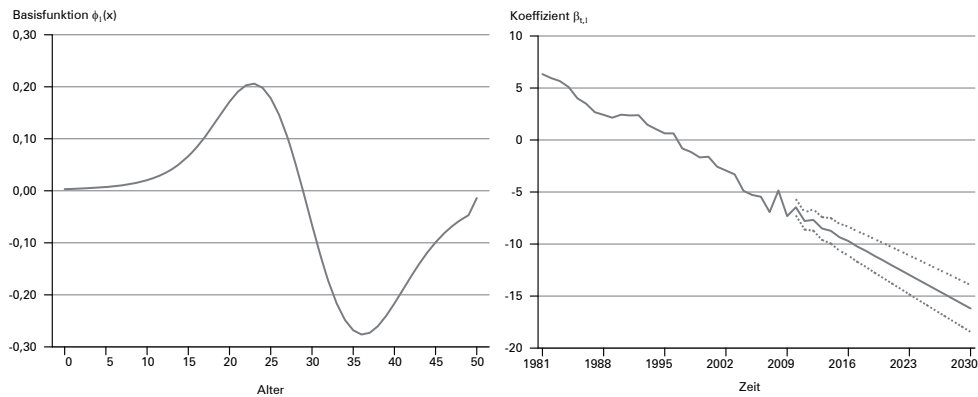
4.3 Die demographischen Komponenten: Modellierung und Prognose

Die geglätteten Daten der Fertilitätsziffer bestehen aus 29 Kurven für die Jahre 1981 bis 2009. Für diesen Zeitraum wurde ein Modell mit $K=6$ Basisfunktionen geschätzt. *Hyndman* und *Booth* (2008: 327) haben gezeigt, dass die Ergebnisse der Methode unabhängig von der Anzahl der Basisfunktionen sind und $K=6$ hinreichend groß ist. Die Basisfunktionen erfassen 82,7 %, 11,3 %, 3,3 %, 1,9 %, 0,6 % und 0,1 % oder 99,9 % der vom Mittelwert abweichenden Variation. Der Koeffizient der ersten Basisfunktion (Abb. 7) zeigt einen Rückgang der Fertilitätsziffer ($\beta_{t,1}$ fällt über den Beobachtungszeitraum), der bei Frauen zwischen 20 und 25 Jahren am stärksten wirkt (höchste positive Funktionswerte der Basisfunktion φ_1). Die negativen Werte der Basisfunktion für das Alter zwischen 35 und 40 Jahren zeigen, dass Geburten in diesem Bereich zunehmen: Die erste Basisfunktion beschreibt somit den gesellschaftlichen Trend, Geburten auf einen späteren Zeitpunkt zu verlagern. Die übrigen Basisfunktionen erfassen insgesamt einen zu geringen Anteil der Varianz der geglätteten Funktion zur Darstellung der altersspezifischen Fertilitätsziffern für den Zeitraum von 1981 bis 2009 und eignen sich daher nicht für eine inhaltliche Interpretation.⁶ Sie wirken systematischen statistischen Verzerrungen, die auf Amplitudenrauschen und Phasenverschiebungen zurückgehen, entgegen.

Die Prognose der Fertilitätsziffer erfolgt für die Jahre 2010 bis 2030 (Abb. 8). Die Maxima der altersspezifischen Fertilitätsziffern steigen über den gesamten Zeitraum an und die einzelnen Kurven verschieben sich im Lauf der Zeit in höhere Altersgruppen. Das mittlere Alter der Mütter erhöht sich von 29,85 (2009) auf 31,36 Jahre (2030) und liegt im 80 %-Prognoseintervall [31,06;31,65]. Die zusammengefasste Geburtenziffer (TFR) nimmt über den Prognosehorizont stetig zu. Abbildung 9 enthält neben den Punktschätzern auch die 80 %-Prognoseintervalle. Die mittlere TFR

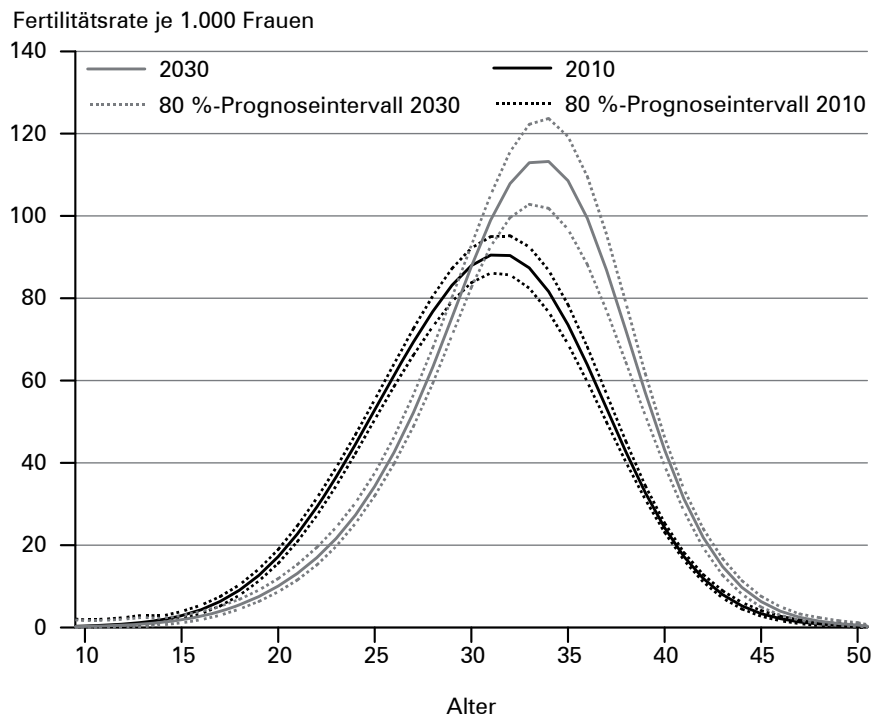
⁶ Alle Basisfunktionen der Modelle für Fertilität, Mortalität und Nettomigration sind im Anhang abgebildet.

Abb. 7: Die erste Basisfunktion des Fertilitäts-Modells inklusive der Prognose des Koeffizienten $\beta_{t,1}$ von 2010 bis 2030 und dem 80 %-Prognoseintervall (gepunktete Linien)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 8: Prognose der Fertilitätsziffer je 1000 Frauen für 2010 (schwarz) und 2030 (grau) mit den 80 %-Prognoseintervallen (gepunktete Linien)



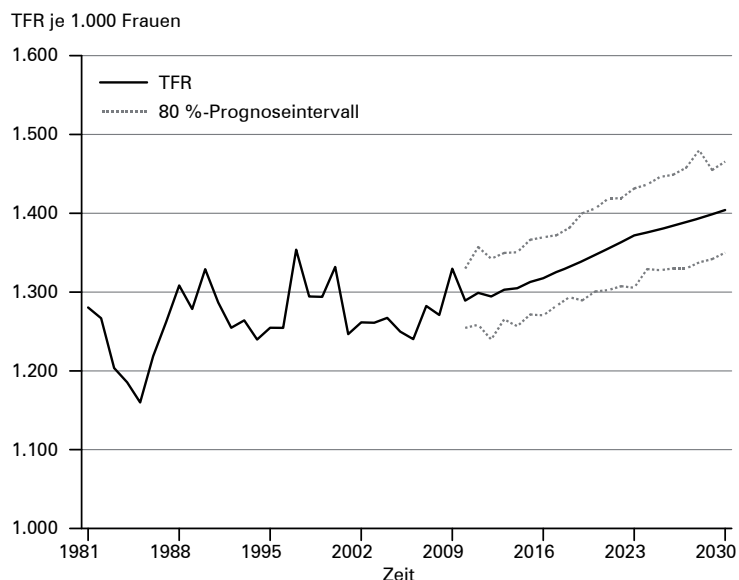
Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

gleicht sich somit den Erwartungen über die Entwicklung der bundesweiten TFR mit 1,4 Geburten pro Frau an (*Statistisches Bundesamt 2009: 27*).

Heckman und Walker (1990) führen eine hohe Korrelation zwischen dem Gebärverhalten und dem Einkommen der Frauen als Grund für die Verlagerung von Geburten in spätere Lebensabschnitte an. Diese ökonomische Interpretation greift allerdings zu kurz. *Ott et al. (2006)* skizzieren ein Drei-Phasen-Modell, das Kinder erst nach dem Erreichen bestimmter beruflicher Ziele in der Lebensplanung berücksichtigt. *Gustaffson (2001)* begründet die Verlagerung durch Nachteile der Mutterschaft auf die Karriere potentieller Mütter. *Börsch-Supan und Wilke (2009)* gehen davon aus, dass durch den sich abzeichnenden Rückgang der Menschen im erwerbsfähigen Alter die Erwerbsbeteiligung von Frauen ansteigen wird und sich dem Erwerbsverhalten von Männern angleicht. Da der Verschiebung der Geburten allerdings biologische Grenzen gesetzt sind, müssen die Geburten in späteren Altersjahren verstärkt nachgeholt werden. Diese Effekte finden sich in der Prognose: Die Geburten werden zeitlich verlagert, das durchschnittliche Alter der Mütter bei der Geburt erhöht sich und die Geburtenziffern der höheren Altersklassen nehmen deutlich zu. In Zukunft tragen Frauen in den hohen Altersklassen am stärksten zur modellierten TFR bei (Abb. 9).

Um zu verhindern, dass die Mortalitätsziffern bei der Prognose divergieren, basiert die Modellierung auf dem kohärenten Modell funktionaler Daten, das sowohl

Abb. 9: Entwicklung der TFR je 1000 Frauen von 1981 bis 2030 mit dem 80 %-Prognoseintervall

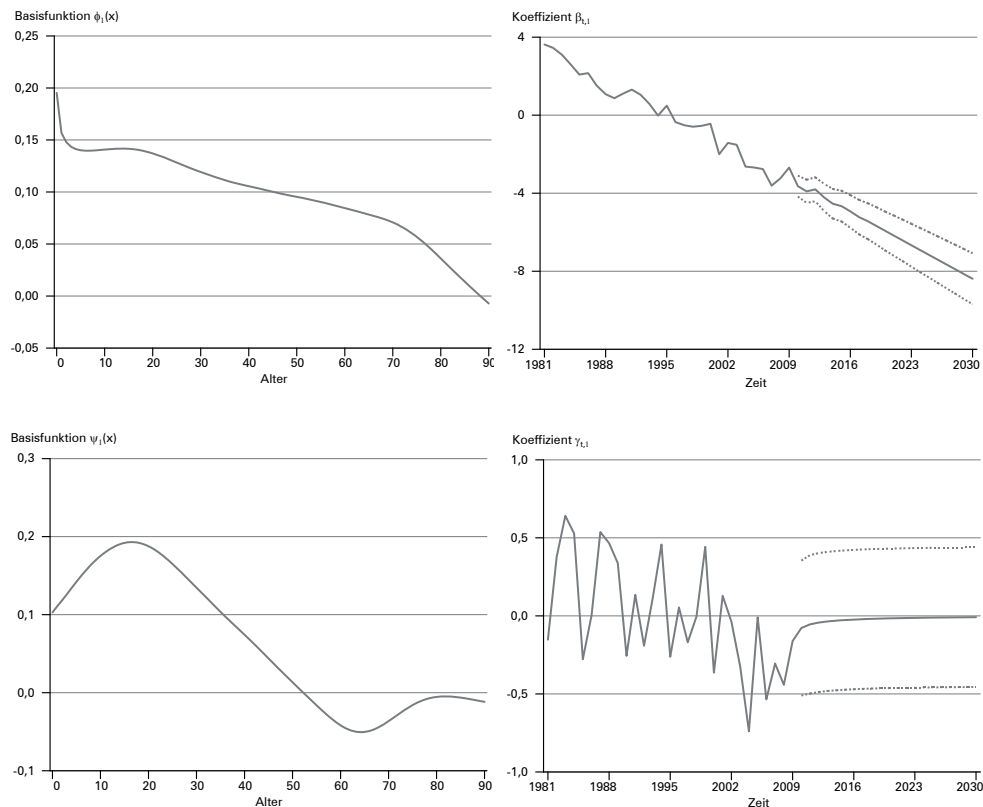


Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

für die Modelle des Produkts als auch für das Verhältnis sechs Basisfunktionen enthält. Die Basisfunktionen des geometrischen Mittels beider Geschlechter erfassen 96,6 %, 1,5 %, 1,0 %, 0,5 %, 0,2 % und 0,1 % der Variation der den Mortalitätsziffern zugrunde gelegten glatten Funktionen. Der Koeffizient $\beta_{t,1}$ der dominierenden ersten Basisfunktion (Abb. 10) zeigt einen Rückgang der Mortalitätsziffern, der sich am stärksten in der Säuglingssterblichkeit (< 1 Jahr) niederschlägt und sich im Alter abschwächt. Die Basisfunktionen des Verhältnis-Modells erfassen 60,2 %, 17,6 %, 10,2 %, 6,8 %, 3,0 % und 1,8 % der Variation. Die erste Basisfunktion zeigt, dass die größten Abweichungen in den Mortalitätsziffern zwischen Männern und Frauen bei etwa 20 Jahren liegen. Diese Variation unterliegt allerdings über die Zeit starken Fluktuationen.

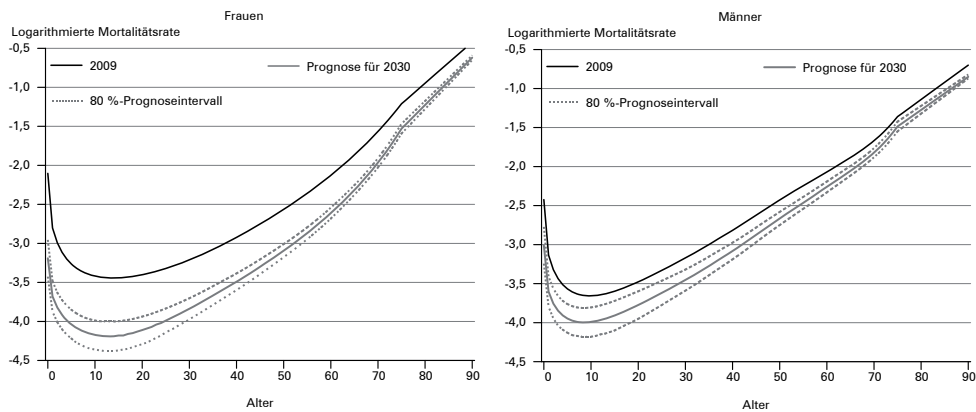
Die Prognose der Mortalität verweist auf den Trend anhaltend sinkender Ziffern, der sich allerdings im Zeitverlauf etwas verlangsamt (Abb. 11). Die Lebenserwartung

Abb. 10: Die ersten Basisfunktionen des Mortalitäts-Modells inklusive der Prognose der Koeffizienten $\beta_{t,1}$ und $\gamma_{t,1}$ von 2010 bis 2030 und dem 80 %-Prognoseintervall (gepunktete Linien)



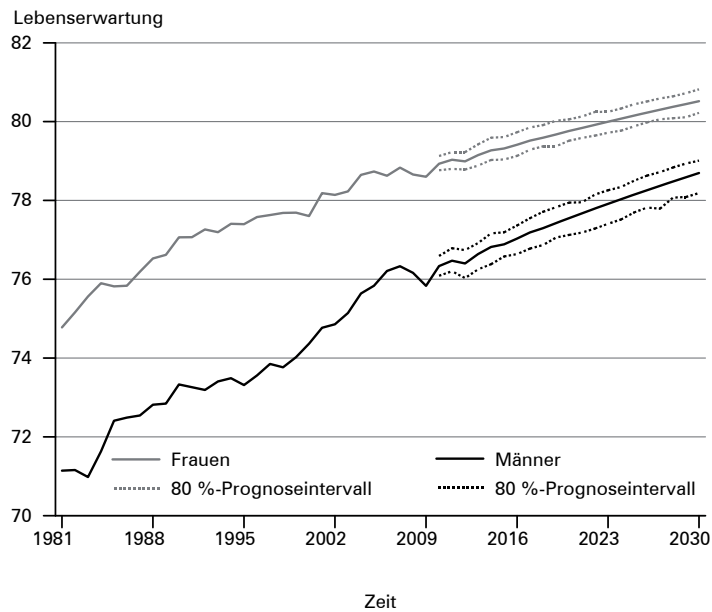
Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 11: Alters- und Geschlechtsspezifische Mortalitätsziffern von 2009 (schwarz) und die Prognose für 2030 (grau) mit den 80 %-Prognoseintervallen (gepunktete Linien)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 12: Entwicklung der Lebenserwartung von Männern (schwarz) und Frauen (grau) von 1981-2030 und den 80 %-Prognoseintervallen (gepunktete Linien)

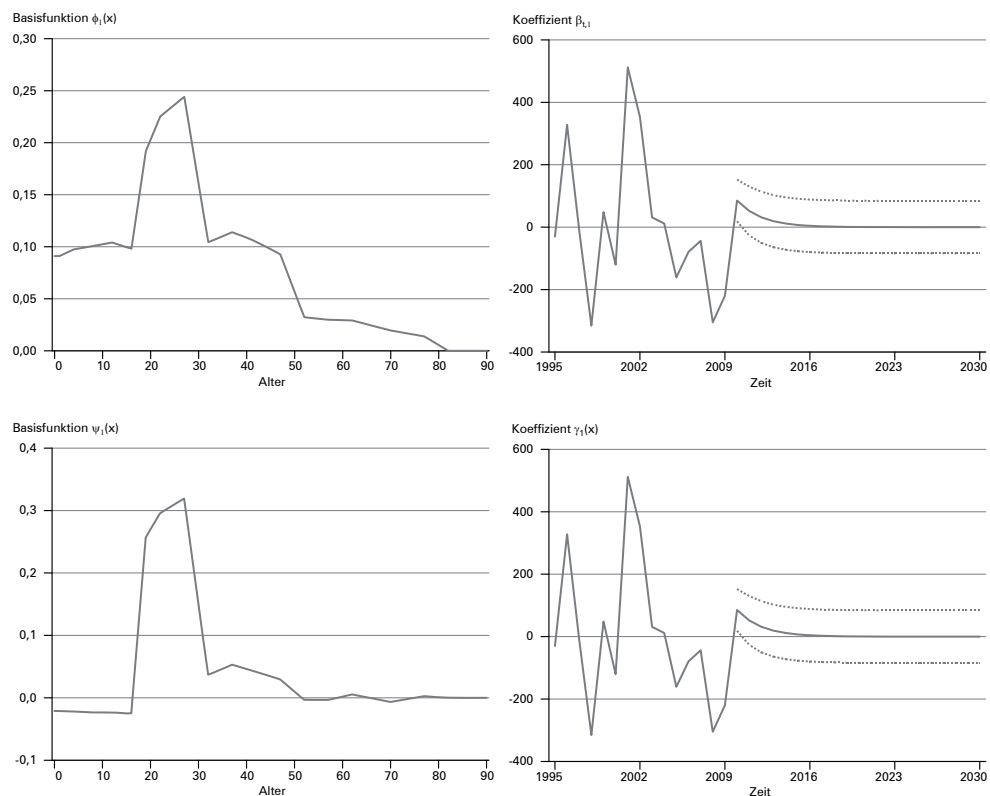


Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

(Abb. 12) von Frauen steigt bis 2030 auf 80,55 Jahre und liegt im 80 %-Prognoseintervall [80.26;80.82], die der Männer auf 78,75 Jahre mit dem 80 %-Prognoseintervall [78,47;79,06]. *Carnes und Olshansky (2007: 377)* nennen die Fettleibigkeit bei Kindern und Jugendlichen sowie neue Infektionskrankheiten als mögliche Quellen, die diesen Trend abschwächen oder gar umkehren können. Diese möglichen zukünftigen Einflüsse können jedoch in einer auf Zeitreihenmodellen basierenden Prognose nicht erfasst werden.

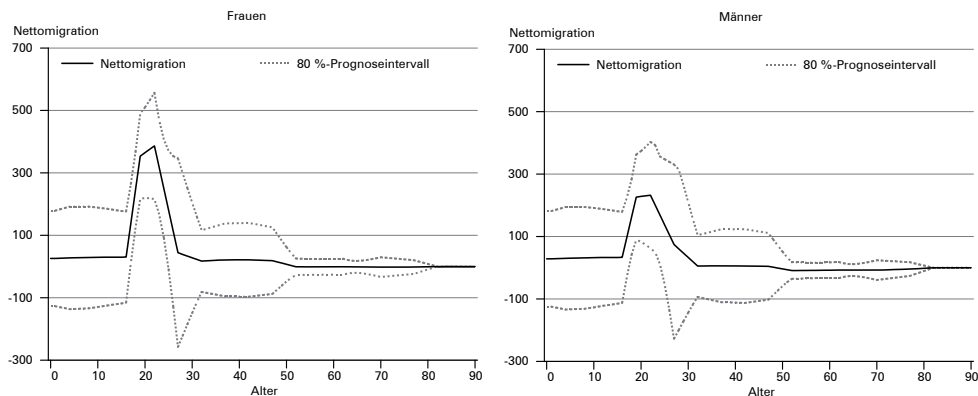
Die aggregierte Nettomigration (Abb. 5) zeigt, dass die Werte für die beiden Geschlechter vergleichbar verlaufen, was eine Modellierung durch das kohärente Modell rechtfertigt. Die Zeitreihe vor 1995 unterliegt einer großen Dynamik, die sich negativ auf die Modellierung und die anschließende Prognose auswirkte. Aus diesem Grund gehen nur die Werte zwischen 1995 und 2009 in die Modellierung ein. Die hohe männliche und weibliche Nettomigration zwischen 18 und 25 Jahren stellt den dominierenden Effekt im Modell dar. Die erste der sechs Basisfunktionen

Abb. 13: Die ersten Basisfunktionen des Migrations-Modells inklusive der Prognose der Koeffizienten $\beta_{t,1}$ und $\gamma_{t,1}$ bis 2030 und dem 80 %-Prognoseintervall (gepunktete Linien)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Abb. 14: Prognose der Altersverteilung der Nettomigration von Männern und Frauen für 2030 mit den 80 %-Prognoseintervallen (gepunktete Linien)



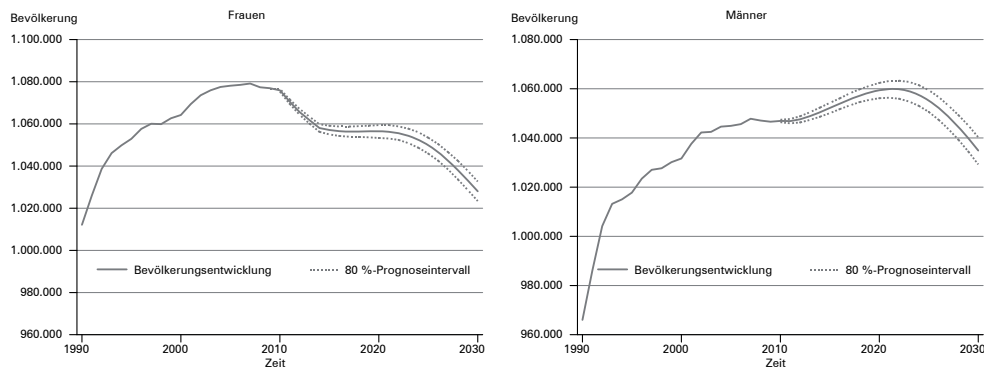
Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

des Produkt-Modells bildet die Veränderung im Zeitverlauf ab (Abb. 13) und erfasst 81,6 % der Variation vom Mittelwert, die restlichen dagegen nur 10,5 %, 5,2 %, 1,4 %, 0,7 % und 0,4 %. Die erste Basisfunktion des Verhältnis-Modells verdeutlicht die geschlechtsspezifischen Unterschiede der Ausbildungsmigranten (Varianzanteil 53,9 %). Der Koeffizient zeigt, dass die männliche Nettomigration zwischen 2000 und 2003 höher war, sich der Trend aber in den Folgejahren abschwächt und sogar umkehrt. Diese Entwicklung setzt sich in der Prognose fort, denn 2030 (Abb. 14) sind im Gegensatz zu 2009 (Abb. 6) die Werte der Frauen im Altersbereich zwischen 18 und 25 Jahren deutlich höher als die der Männer (Abb. 14). Im Modell bleibt die Metropolregion Rhein-Neckar ein attraktiver Ausbildungs- oder Studienstandort für junge Menschen. Jedoch stellen die Prognosen der Nettomigration durch die weiten Prognoseintervalle für Menschen unter 50 Jahre die größte Quelle für Unsicherheit über die zukünftige Bevölkerungsentwicklung dar.

4.4 Die Bevölkerungsentwicklung der Metropolregion Rhein-Neckar

Seit den 1990er Jahren steigt die männliche und weibliche Bevölkerung der Metropolregion Rhein-Neckar aufgrund von Wanderungsgewinnen durch die politischen Entwicklungen in Deutschland (Wiedervereinigung) und Europa (Zuwanderung von Spätaussiedlern) stetig an (*Gans/Schmitz-Veltin* 2006: 316). Die Bevölkerungsentwicklung für die Jahre 2010 bis 2030 berechnet sich aus der Anwendung des in Abschnitt 4.1 beschriebenen Algorithmus auf Basis der Kohorten-Komponenten-Methode und den aus den demographischen Komponenten ermittelten Geburten, Sterbefällen und Migranten. Die männliche Bevölkerung steigt bis 2022 leicht an (Abb. 15) und sinkt aber gemessen am Median der Simulation bis 2030 auf 1,0349 Millionen Einwohner, was einem Rückgang von 2,31 % für den Prognosehorizont entspricht, und liegt im 80 %-Prognoseintervall [1,0292;1,0403]. Die Anzahl der

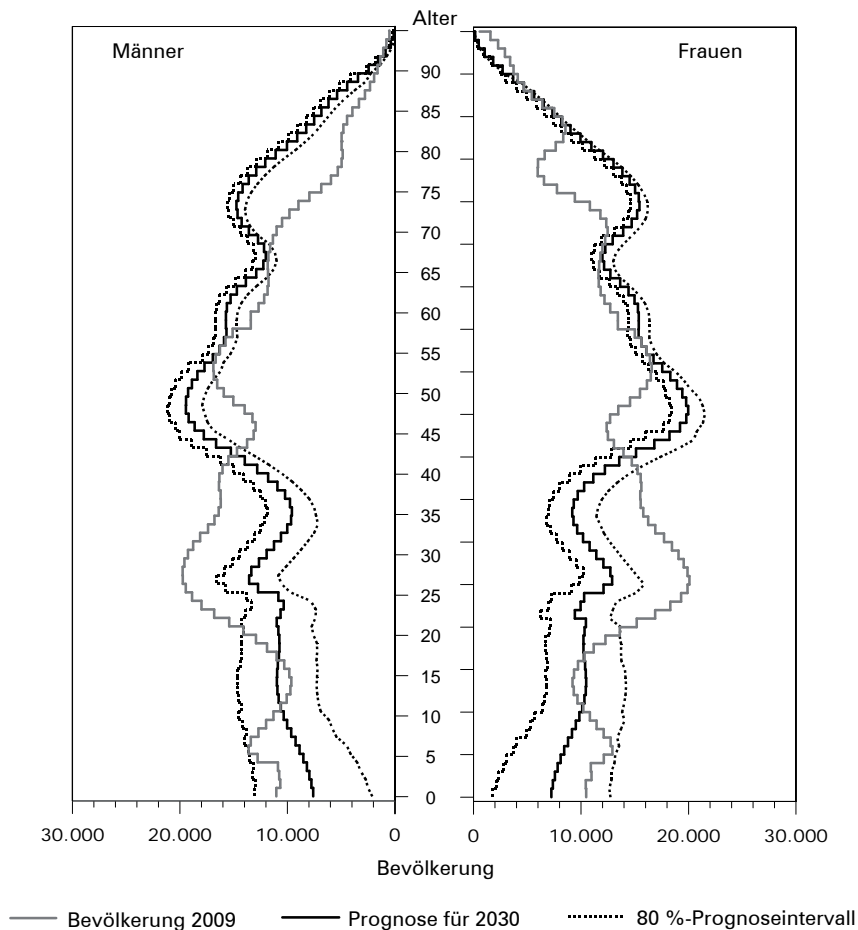
Abb. 15: Entwicklung der männlichen und weiblichen Bevölkerung zwischen 2010 und 2030 mit den 80 %-Prognoseintervallen (gepunktete Linien)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

Frauen fällt bis 2030 um 4,45 Prozent auf 1,0280 Millionen Einwohner (80 %-Prognoseintervall [1,0234;1,0327]). Die weibliche Bevölkerung nimmt demnach etwas stärker ab als die männliche. Die Bevölkerungspyramide in Abbildung 16 zeigt den erwarteten Altersaufbau für die Jahre 2009 als letztes Jahr der Zeitreihendaten und die Prognose für 2030 mit den 80 %-Prognoseintervallen. Den größten Anteil an der Gesamtbevölkerung stellen 2009 die Altersgruppen zwischen 20 und 30 Jahren. Die Fertilitätsziffern liegen deutlich unterhalb des Niveaus zur Bestandserhaltung der Bevölkerung und auch die positive Nettomigration reicht nicht aus, um die Sterbeüberschüsse auszugleichen. Damit zeichnet sich ein deutlicher Rückgang der Menschen unter 30 Jahren ab. Die Zahl der Menschen im erwerbsfähigen Alter unter 30 Jahren verringert sich massiv, während die Gruppe der 30- bis unter 65-Jährigen etwa das Niveau von 2009 hält. Dieser Effekt entsteht im Wesentlichen durch die Verlagerung der zahlenmäßig größten Jahrgänge aus 2009 auf die älteren Altersgruppen in 2030. Das Prognoseintervall ist im Altersbereich unter 20 Jahre am größten, da die Anzahl dieser Menschen aus der Prognose der Fertilität hervorgeht, während die Menschen im Jahr 2030 im Altersbereich über 20 Jahre zum ersten Prognosejahr (2010) bereits geboren waren und somit keine Quelle zusätzlicher Unsicherheit darstellen. Die Weite der Konfidenzintervalle bestimmt sich darüber hinaus durch die Variation der Nettomigration, die sich ab 50 Jahren deutlich stabilisiert (vgl. Abb. 14), wodurch die Bevölkerung ab 50 Jahren mit einer hohen Genauigkeit vorhergesagt werden kann.

Abb. 16: Männliche und weibliche Bevölkerung nach Altersjahren 2009 (grau) und das 0,5-Quantil der Simulation für 2030 (schwarz) mit den 80 %-Prognoseintervallen (gepunktete Linien)



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz.

5 Schlussfolgerungen

Die demographische Zukunft einer Region ist unsicher. Die von den offiziellen Statistikstellen häufig für Vorausberechnungen verwendeten deterministischen Modelle weisen einen „wenn... dann...“-Charakter auf, die demographischen Determinanten entwickeln sich über vorgegebene Annahmen in verschiedenen Szenarien. Diese sollen die Bandbreite der möglichen Entwicklungen abbilden, allerdings ohne Aussagen über deren Eintrittswahrscheinlichkeiten zu treffen. Als problemorien-

tiert erweist sich der stochastische Ansatz, der durch Zeitreihenmodelle neben der zukünftigen Bevölkerung auch ein Prognoseintervall für die demographischen Parameter berechnet. Diese Herangehensweise erfordert jedoch lange Zeitreihen und nach Altersjahren gegliederte Daten über alle Komponenten der demographischen Grundgleichung. Derartige Daten sind auf regionaler Ebene allerdings häufig gar nicht, nur in groben Altersgruppen oder in unzureichender Struktur verfügbar. Darüber hinaus fassen die verschiedenen statistischen Landesämter die Merkmale zu unterschiedlichen Altersgruppen zusammen, eine Aggregation auf eine bundeslandübergreifende Gesamtregion wie die Metropolregion Rhein-Neckar, die aus Stadt- und Landkreisen aus Baden-Württemberg, Hessen und Rheinland-Pfalz besteht, ist auf einer solchen Basis faktisch unmöglich. Für diese Probleme erweist sich das Paradigma funktionaler Daten als äußerst hilfreich, denn per Annahme folgen altersgruppierte Sekundärdaten einem funktionalen Zusammenhang, der mit den vorgestellten FDA-Modellen geschätzt werden kann. Auf diese Weise entstehen Datensätze, die nach Altersjahren gegliedert sind und daher zu grenzübergreifenden Funktionalregionen aggregiert werden können. So ermöglichen das Paradigma funktionaler Daten und die darauf aufbauenden Modellierungen insbesondere auf regionaler Ebene sekundärstatistische Datenlücken zu schließen und verlässlichere Aussagen über die demographische Entwicklung zu treffen als mit Hilfe deterministischer Ansätze. Ohne die Ansätze aus Kapitel 3 sind stochastische Modellierungen auf regionaler Ebene nicht durchführbar. Durch die Prognoseintervalle auf Basis der Zeitreihenmodelle entsteht darüber hinaus eine wertvolle Zusatzinformation für regionale Planer gegenüber deterministischen Vorausberechnungen. Denn gerade auf regionaler Ebene besteht ein großer Bedarf an Informationen über die Bevölkerungsentwicklung als Grundlage zur Abschätzung der Konsequenzen des demographischen Wandels, beispielsweise auf das Arbeitsangebot und insbesondere auf die Anzahl der Erwerbspersonen. Darüber hinaus dienen regionale Bevölkerungsprognosen als Grundlage zur Abschätzung des zukünftigen Wohnflächenbedarfs.

Die Anwendung der Modelle für funktionale Daten auf die Metropolregion Rhein-Neckar verdeutlicht die Konsequenzen des demographischen Wandels für den siebtgrößten deutschen Wirtschaftsraum und stellt eine Möglichkeit zur Realisierung für stochastische Bevölkerungsprognosen auf regionaler Ebene dar. Die Bevölkerung nimmt bis 2030 zwar nur geringfügig ab, allerdings steht einem erheblichen Rückgang der jungen Menschen im erwerbsfähigen Alter unter 30 Jahren ein beträchtlicher Anstieg der Menschen in den oberen Altersgruppen und insbesondere im Rentenalter gegenüber. Diese Knappheit kann durch Zuwanderung abgeschwächt werden. Grundvoraussetzung hierfür ist, die Stärken der eigenen Region zu erkennen und Anreize für junge Menschen zu schaffen, in die Region zu ziehen und dort nach dem Studium oder der Ausbildung auch zu bleiben. Die Auswirkungen des demographischen Wandels werden sich insbesondere auf dem Arbeitsmarkt bemerkbar machen. Regionen stehen im Wettbewerb um junge und gut ausgebildete Erwerbstätige. Aus diesem „Kampf um kluge Köpfe“ leitet sich zukünftiger Forschungsbedarf ab: Stochastische Bevölkerungsprognosen können als Grundlage für die Abschätzung der Entwicklung des Arbeitsangebotes genutzt werden. Jedoch basieren viele existierende Vorausberechnungen auf deterministischen Mo-

dellen, da unter anderem arbeitsmarktrelevante Größen, wie die Arbeitslosenquote, politischen Einflüssen unterliegen und somit nur schwer von Zeitreihenmodellen erklärt werden können. Hier gilt es Methoden zu entwickeln, die diese Einflüsse berücksichtigen können (beispielsweise bei der Generierung der Annahmen) und die Vorteile beider Ansätze in Einklang bringen. Das Paradigma funktionaler Daten kann hierzu einen wertvollen Beitrag leisten.

Literatur

- Alho, Juha; Spencer, Bruce* 2005: Statistical demography and forecasting. New York: Springer.
- Bähr, Jürgen; Jentsch, Christoph; Kuls, Wolfgang* 1992: Bevölkerungsgeographie. Berlin: De Gruyter.
- Birg, Herwig* 2005: Grundkurs Demographie – Dritte Lektion: Was taugen Prognosen? In: Frankfurter Allgemeine Zeitung 46, 24.02.2005: 41.
- Booth, Heather* 2006: Demographic forecasting: 1980 to 2005 in review. In: International Journal of Forecasting 22,3: 547–581 [doi:10.1016/j.ijforecast.2006.04.001].
- Börsch-Supan, Axel* 2007: Gesamtwirtschaftliche Folgen des demographischen Wandels. In: Geographische Rundschau 2: 48-52.
- Börsch-Supan, Axel; Wilke, Christina Benita* 2009: Zur mittel- und langfristigen Entwicklung der Erwerbstätigkeit in Deutschland. In: Zeitschrift für Arbeitsmarktforschung (ZAF) 42,1: 29-48 [doi: 10.1007/s12651-009-0006-x].
- Brillinger, David* 1986: The Natural variability of vital rates and associated statistics. In: Biometrics 42: 693-734.
- Bundesamt für Bauwesen, Raumordnung und Städtebau* 1995: Raumordnungspolitischer Handlungsrahmen. Bonn.
- Büttner, Thiess; Ebertz, Alexander* 2009: Quality of Life in the Regions - Results for German Counties. In: Annals of Regional Science 43,1: 89-112 [doi: 10.1007/s00168-007-0204-9].
- Carnes, Bruce; Olshansky, Jay* 2007: A Realistic View of Aging, Mortality and Future Longevity. In: Population and Development Review 33,2: 367-381.
- Chambers, Elizabeth et al.* 1998: The War for Talent. In: McKinsey Quarterly 3. [http://www.executivesondemand.net/management sourcing/images/stories/artigos_pdf/gestao/The_war_for_talent.pdf, 09.06.2011].
- De Jong, Gordon* 2000: Expectations, gender, and norms in migration decision-making. In: Population Studies 54: 307–319.
- Erbas, Bircan; Hyndman, Rob; Gertig, Dorota* 2007: Forecasting age-specific breast cancer mortality using functional data models. In: Statistics in Medicine 26,2: 458–470 [doi: 10.1002/sim.2306].
- Fuchs, Johann; Dörfler, Katrin* 2005a: Projektion des Erwerbspotenzials bis 2050 – Annahmen und Datengrundlage. In: IAB Forschungsbericht 20/2005. Nürnberg: IAB.
- Fuchs, Johann; Dörfler, Katrin* 2005b: Projektion des Arbeitsangebots bis 2050 – Demographische Effekte sind nicht mehr zu bremsen. In: IAB Kurzbericht 11/2005. Nürnberg: IAB.

- Gans, Paul; Schmitz-Veltin, Ansgar* 2006: Die Auswirkungen des demographischen Wandels auf den Agglomerationsraum Rhein-Neckar – Ausgewählte regionale Entwicklungsbereiche vor dem Hintergrund sich ändernder Bevölkerungs- und Wirtschaftsstrukturen. In: *Gans, Paul; Schmitz-Veltin, Ansgar* (Hrsg.): Demographische Trends in Deutschland. Folgen für Städte und Regionen. Räumliche Konsequenzen des demographischen Wandels, Teil 6. Forschungs- und Sitzungsberichte der ARL 226. Hannover: 311-335.
- Gans, Paul; Schmitz-Veltin, Ansgar; West, Christina* 2009: Bevölkerungsgeographie. Braunschweig: Westermann.
- Gustafsson, Siv* 2001: Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe. In: *Population Economics* 14: 225-247.
- He, Xuming; Ng, Pin* 1999: COBS: qualitatively constrained smoothing via linear programming. In: *Computational Statistics* 14: 315–337.
- Heckman, James; Walker, James* 1990: The Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data. In: *Econometrica* 58,6: 1411-1441.
- Höhn, Charlotte et al.* 1987: Mehrsprachiges demographisches Wörterbuch. Deutschsprachige Fassung. Schriftenreihe des Bundesinstituts für Bevölkerungsforschung, Sonderband 16. Wiesbaden/Boppard: Boldt.
- Hyndman, Rob; Booth, Heather* 2008: Stochastic population forecasts using functional data models for mortality, fertility and migration. In: *International Journal of Forecasting* 24,3: 323–342 [doi: 10.1016/j.ijforecast.2008.02.009].
- Hyndman, Rob; Ullah, Shahid* 2007: Robust forecasting of mortality and fertility rates: a functional data approach. In: *Computational Statistics & Data Analysis* 51: 4942–4956 [doi: 10.1016/j.csda.2006.07.028].
- Hyndman, Rob; Booth, Heather; Yasmeen, Farah* 2011: Coherent mortality forecasting: the product-ratio method with functional time series models. EBS Working Paper [http://robjhyndman.com/papers/coherentfdm].
- Keilman, Nico; Pham, Dinh Quang; Hetland, Arve* 2002: Why population forecasts should be probabilistic – Illustrated by the case of Norway. In: *Demographic Research* 6: 409–453 [doi: 10.4054/DemRes.2002.6.15].
- Keyfitz, Nathan* 1972: On Future Population. In: *Journal of the American Statistical Association* 67: 347-363.
- Keyfitz, Nathan* 1981: The Limits of Population Forecasting. In: *Population and Development Review* 7,4: 579-593.
- Klein, Torsten* 2008: Kampf um Talente – Das Kind ist (fast) in den Brunnen gefallen. In: *HR Performance* 6: 44-47.
- Lee, Ronald* 1993: Modeling and forecasting the time series of U.S. fertility: age distribution, range, and ultimate level. In: *International Journal of Forecasting* 9: 187–202.
- Lee, Ronald* 1999: Probabilistic approaches to population forecasting. In: *Lutz, Wolfgang; Vaupel, James; Ahlburg, Dennis* (Hrsg.): *Frontiers of population forecasting. A supplement to Population and Development Review* 24: 156 –190 [doi: 10.2307/2808055].
- Lee, Ronald* 2000: The Lee-Carter method for forecasting mortality, with various extensions and applications. In: *North American Actuarial Journal* 4,1: 80–92.
- Lee, Ronald; Carter, Lawrence* 1992: Modeling and forecasting US mortality. In: *Journal of the American Statistical Association* 87,419: 659–671.

- Lee, Ronald; Nault, Francois* 1993: Modeling and forecasting provincial mortality in Canada. Paper presented at the World Congress of the IUSSP, Montreal, Canada.
- Lee, Ronald; Tuljapurkar, Shripad* 1994: Stochastic Forecasts for the United States: Beyond High, Medium, and Low. In: *Journal of American Statistical Association* 89,428: 1175-1189.
- Li, Nan; Lee, Ronald* 2005: Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. In: *Demography* 42,3: 575–594 [doi: 10.1353/dem.2005.0021].
- Lipps, Oliver; Betz, Frank* 2003: Stochastische Bevölkerungsprognose für West- und Ostdeutschland. MEA Discussion Paper 41. Mannheim: MEA.
- Lipps, Oliver; Betz, Frank* 2004: Stochastic Population Projection for Germany – based on the Quadratic Spline approach to modelling age specific fertility rates. MEA Discussion Paper 59. Mannheim: MEA.
- Lowack, Wolf-Rainer* 2007: Die Europäische Metropolregion Rhein-Neckar – Auftrag, Vision und Umsetzung. In: *UWF – UmweltWirtschaftsForum* 15,3: 130-135 [doi: 10.1007/s00550-007-0042-3].
- Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei* 1996: World population scenarios for the 21st century. In: *Lutz, Wolfgang* (Hrsg.): *The future population of the world: What can we assume today?* London: Earthscan: 361-396.
- Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei* 1998a: Expert-based probabilistic population projections. In: *Population and Development Review* 24, Supplement: *Frontiers of Population Forecasting*: 139-155 [doi: 10.2307/2808054].
- Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei* 1998b: An expert-based framework for probabilistic national population projections: The example of Austria. In: *European Journal of Population* 14: 1-17.
- Lutz, Wolfgang; Sanderson, Warren; Scherbov, Sergei* 2001: The end of world population growth. In: *Nature* 412: 543-545.
- Metropolregion Rhein-Neckar* 2011: Jahresbericht 2010 zur gemeinschaftlichen Regionalentwicklung. In: www.m-r-n.de 9.2.2012. URL: http://www.m-r-n.com/fileadmin/PDF-Downloads/MRN_Jahresbericht_2010_01.pdf.
- Moore, Andrew; Schneider, Jeff; Deng, Kang* 1997: Efficient Locally Weighted Polynomial Regression Predictions. In: *Kaufmann, Morgan* (Hrsg.): *Proceedings of the Fourteenth International Conference on Machine Learning*. San Francisco, CA, USA: 236-244.
- O’Neill, Brian et al.* 2001: A Guide to Global Population Projections. In: *Demographic Research* 4: 203-288 [doi: 10.4054/DemRes.2001.4.8].
- Ott, Thomas; Schmitz-Veltin, Ansgar; West, Christina* 2006: Kinderwunsch – immer später und oft unerfüllt. In: *Institut für Länderkunde* (Hrsg.): *Nationalatlas Bundesrepublik Deutschland*, Bd. 12, *Leben in Deutschland*. Heidelberg/Berlin: Spectrum: 48-49.
- Pflaumer, Peter* 1988: Confidence Intervals for Population Projections based on Monte Carlo Methods. In: *International Journal of Forecasting* 4: 135-142.
- Ramsay, Jim* 2008: *Functional Data Analysis*. URL: http://gbi.agrsci.dk/~shd/public/FDA2008/FDA_Sage.pdf, 31.05.2011.
- Ramsay, Jim; Silverman, Bernard* 2001: *Functional Data Analysis*. In: *Smelser, Neil; Balmes, Paul* (Hrsg.): *International Encyclopedia of the Social and Behavioral Sciences*. Amsterdam: Elsevier: 5822-5828.
- Rogers, Andrei* 1985: *Regional Population Projection Models*. Beverly Hills: Sage.

- Sanderson, Warren* 1995: Predictability, complexity, and catastrophe in a collapsible model of population, development, and environmental interactions. In: *Mathematical Population Studies* 5,3: 259-279.
- Schnabel, Sabine; Kistowski, Kristin von; Vaupel, James* 2005: Immer neue Rekorde und kein Ende in Sicht: der Blick in die Zukunft lässt Deutschland grauer aussehen als viele erwarten. In: *Demografische Forschung Aus Erster Hand* 2,2: 3-3.
- Statistisches Bundesamt* 2009: 12. koordinierte Bevölkerungsvorausberechnung – Annahmen und Ergebnisse. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt* 2010: Statistisches Jahrbuch 2010 für die Bundesrepublik Deutschland. Wiesbaden: Statistisches Bundesamt.
- Stoto, Michael* 1983: The accuracy of population projections. In: *Journal of the American Statistical Association* 78: 13–20.
- Tukey, John Wilder* 1977: *Exploratory Data Analysis*. London: Addison-Wesley.
- Whelpton, Pascal* 1928: Population of the United States, 1925 to 1975. In: *American Journal of Sociology* 34: 253-270.
- Wilson, Chris* 2001: On the scale of Global Demographic Convergence 1950-2000. In: *Population and Development Review* 27,1: 155-171 [doi: 10.1111/j.1728-4457.2001.00155.x].
- Wood, Simon* 1994: Obtaining Birth and Mortality Patterns From Structured Population Trajectories. In: *Ecological Monographs* 64,1: 23-44 [doi: 10.2307/2937054].

Eine Übersetzung dieses begutachteten und vom Autor autorisierten deutschen Originaltextes durch das Bundesinstitut für Bevölkerungsforschung ist unter dem Titel „Population Development of the Rhine-Neckar Metropolitan Area: A Stochastic Population Forecast on the Basis of Functional Data Analysis“, DOI 10.4232/10.CPoS-2011-21en bzw. URN urn:nbn:de:bib-cpos-2011-21en4, auf <http://www.comparativepopulationstudies.de> verfügbar.

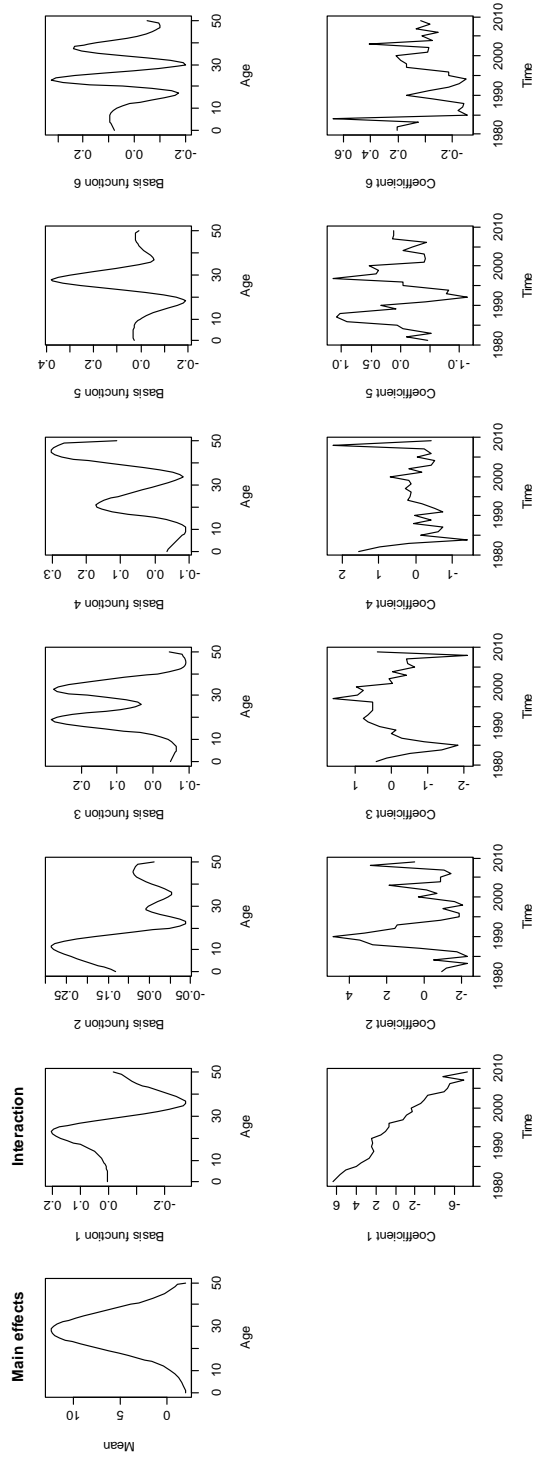
Eingegangen am: 29.08.2011

Angenommen am: 25.04.2012

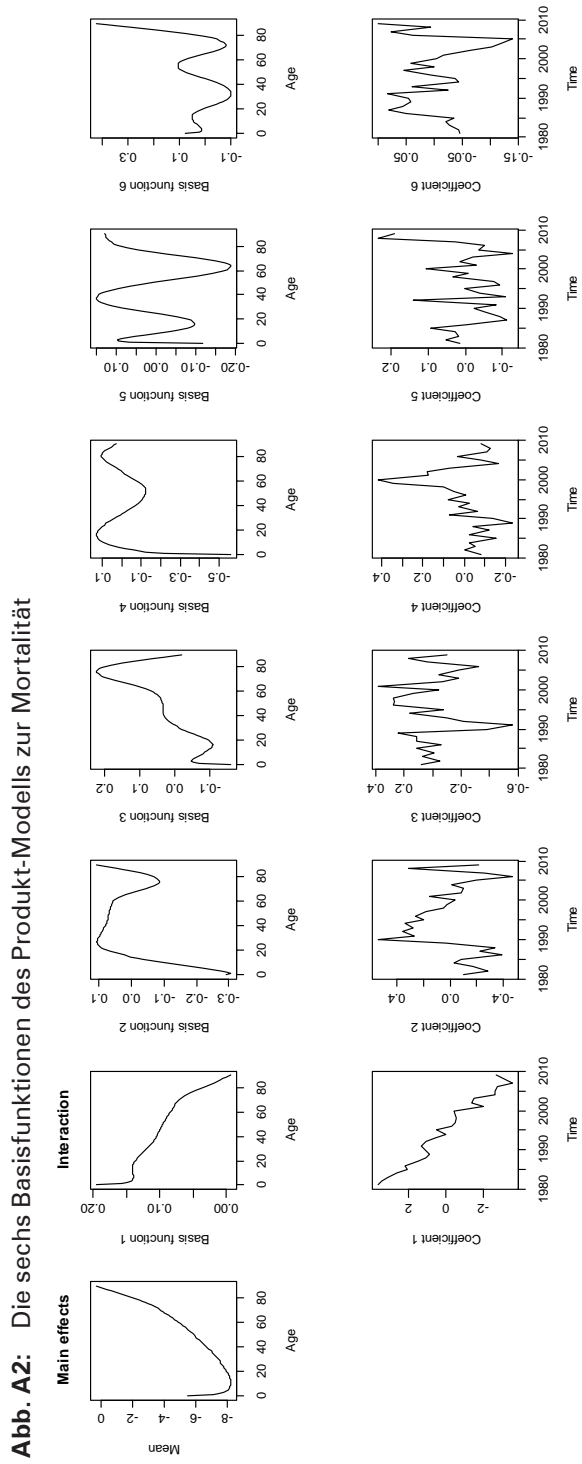
Philipp Deschermeier (✉). Universität Mannheim, Lehrstuhl für Wirtschaftsgeographie, Fakultät für Rechtswissenschaft und Volkswirtschaftslehre. Mannheim, Deutschland.
E-Mail: philipp.deschermeier@uni-mannheim.de
URL: <http://gans.vwl.uni-mannheim.de/1465.0.html>

Anhang

Abb. A1: Die sechs Basisfunktionen des Fertilitätsmodells

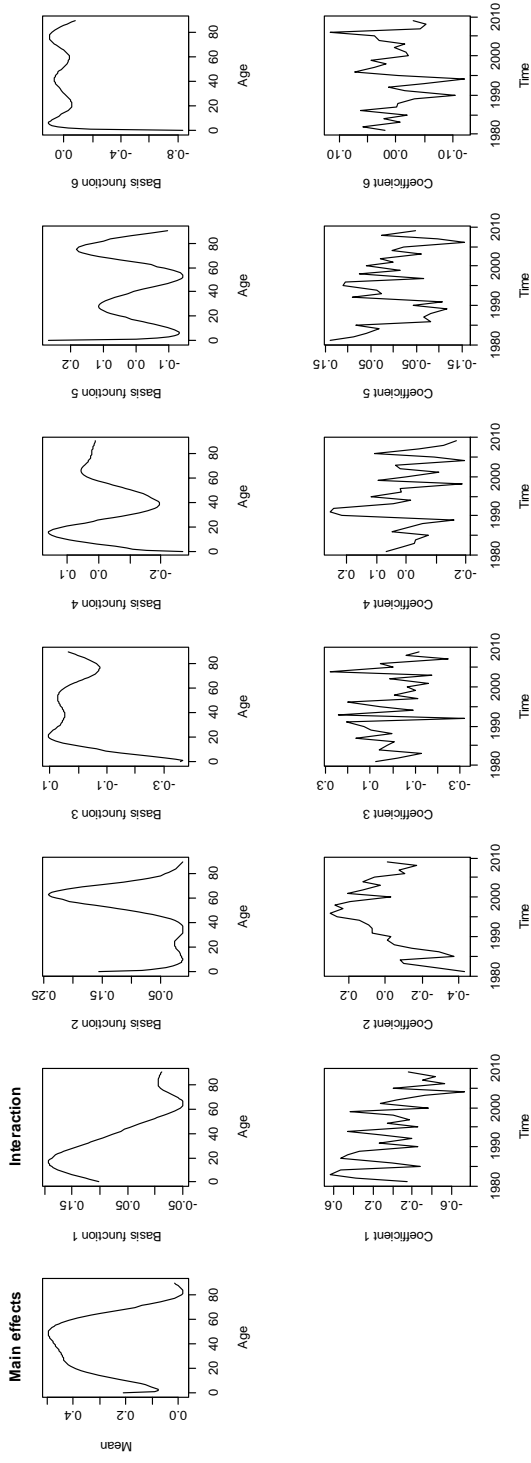


Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz

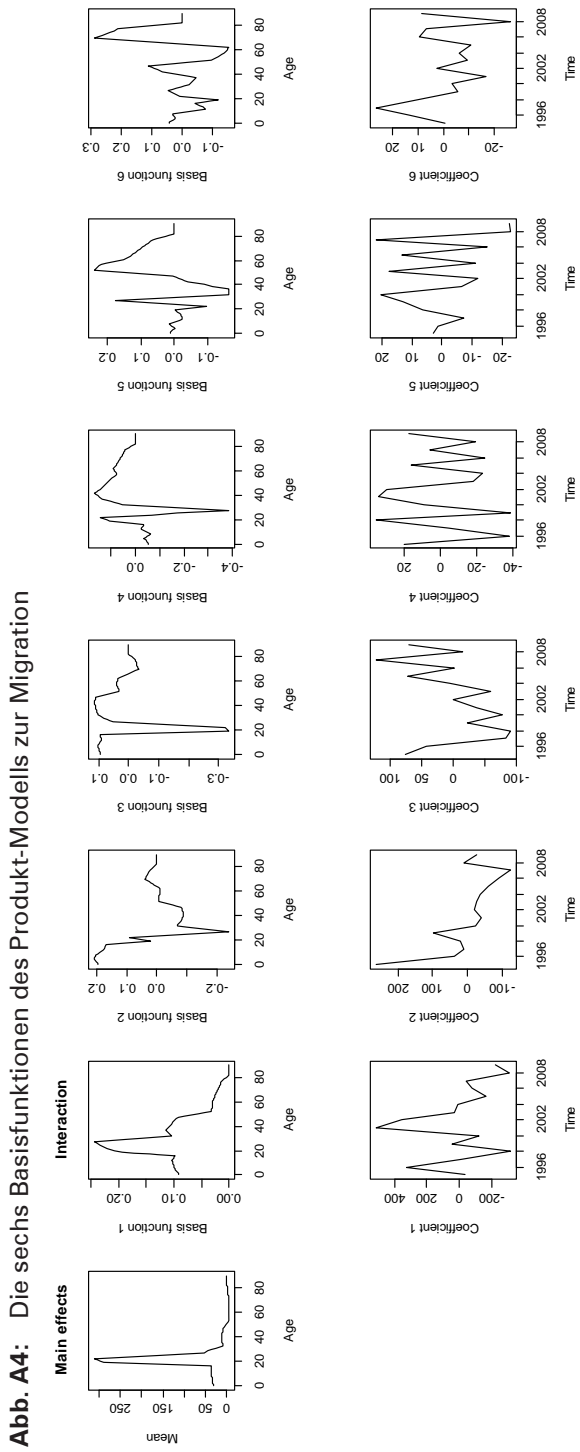


Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz

Abb. A3: Die sechs Basisfunktionen des Verhältnis-Modells zur Mortalität

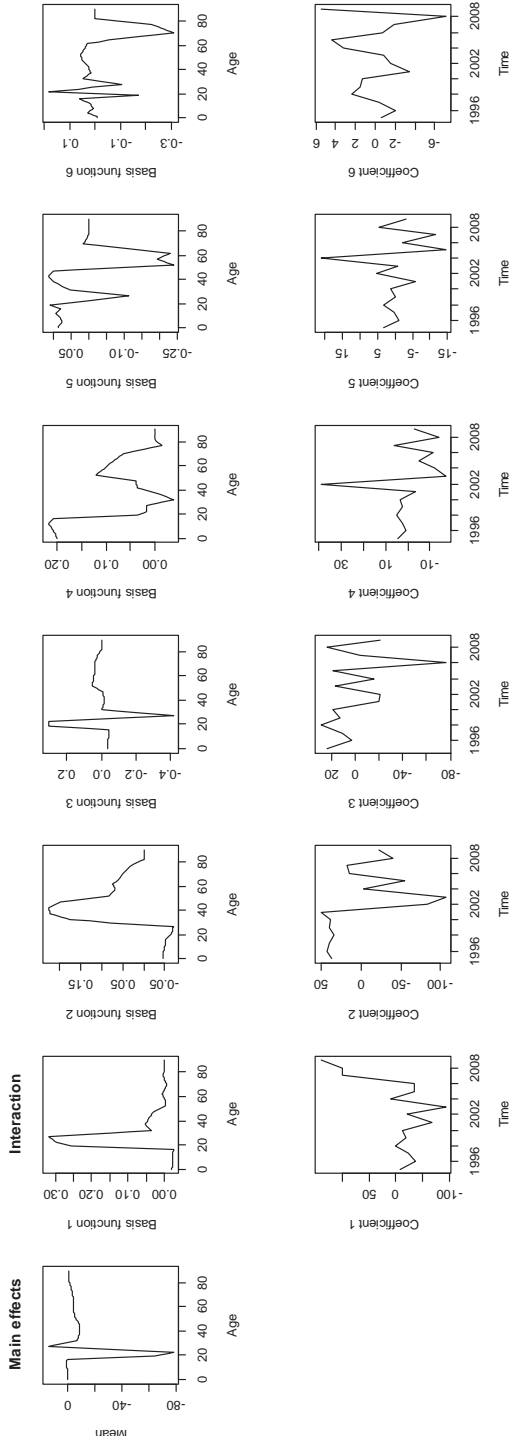


Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz

Abb. A5: Die sechs Basisfunktionen des Verhältnis-Modells zur Migration



Quelle: Eigene Berechnung auf Basis der Daten der Statistischen Landesämter Baden-Württemberg, Hessen und Rheinland-Pfalz

Comparative Population Studies – Zeitschrift für Bevölkerungswissenschaft

www.comparativepopulationstudies.de

ISSN: 1869-8980 (Print) – 1869-8999 (Internet)

Published by / Herausgegeben von

Prof. Dr. Norbert F. Schneider

Federal Institute for Population Research
D-65180 Wiesbaden / Germany

Managing Editor /

Verantwortlicher Redakteur

Frank Swiaczny

Editorial Assistant /

Redaktionsassistent

Katrin Schiefer

Language & Copy Editor (English) /

Lektorat & Übersetzungen (englisch)

Amelie Franke

Copy Editor (German) /

Lektorat (deutsch)

Dr. Evelyn Grünheid

Layout / Satz

Beatriz Feiler-Fuchs

E-mail: cpos@destatis.de

Scientific Advisory Board /

Wissenschaftlicher Beirat

Jürgen Dorbritz (Wiesbaden)

Paul Gans (Mannheim)

Johannes Huinink (Bremen)

Marc Luy (Wien)

Clara H. Mulder (Groningen)

Notburga Ott (Bochum)

Peter Preisendörfer (Mainz)

Board of Reviewers / Gutachterbeirat

Martin Abraham (Erlangen)

Laura Bernardi (Lausanne)

Hansjörg Bucher (Bonn)

Claudia Diehl (Göttingen)

Andreas Diekmann (Zürich)

Gabriele Doblhammer-Reiter (Rostock)

Henriette Engelhardt-Wölfler (Bamberg)

E.-Jürgen Flöthmann (Bielefeld)

Alexia Fürnkranz-Prskawetz (Wien)

Beat Fux (Zürich)

Joshua Goldstein (Rostock)

Karsten Hank (Köln)

Sonja Haug (Regensburg)

Franz-Josef Kemper (Berlin)

Michaela Kreyenfeld (Rostock)

Aart C. Liefbroer (Den Haag)

Kurt Lüscher (Konstanz)

Dimiter Philipov (Wien)

Tomáš Sobotka (Wien)

Heike Trappe (Rostock)