

Zur stochastischen Variabilität von Bevölkerungsprognosen für Regionen bzw. für Bevölkerungsgruppen — Vergleich alternativer Simulationsmodelle am Fallbeispiel der Stadt Werne

HERWIG BIRG und RÜDIGER BROCK

Zusammenfassung

Die Variabilität von Bevölkerungsprognosen beruht auf drei verschiedenen Fehlerquellen, auf Fehlern der Ausgangsdaten, auf den Abweichungen der Prognoseannahmen über Ursache-Wirkungsbeziehungen von den realen Zusammenhängen (Annahmefehlern) sowie auf der grundsätzlich stochastischen Eigenschaft des Bevölkerungsprozesses. In der Praxis haben die Annahmefehler das größte Gewicht, bei regionalen Bevölkerungsprognosen vor allem die Annahmefehler bezüglich der Wanderungen.

Die stochastische Variabilität hängt von der Größe der Region ab (Einwohnerzahl). Für das Fallbeispiel der Stadt Werne (Größenklasse um 30 000 Einwohner) wurde die stochastische Variabilität auf der Basis eines Modells ohne Wanderungen für zwei verschiedene Modelltypen untersucht, und zwar für ein *Einzelsimulationsmodell* und für ein *Regions- bzw. Gruppensimulationsmodell*. Während die Erwartungswerte der prognostizierten bzw. simulierten Bevölkerungszahl — gegliedert nach Alter und Geschlecht — in beiden Modellen weitgehend übereinstimmen, ist ihre stochastische Variabilität, gemessen durch den Variationskoeffizienten der Prognoseergebnisse aus einer großen Zahl von Simulations- bzw. Prognoseläufen, unterschiedlich. Die stochastische Variabilität des Einzelsimulationsmodells ist in dem vorliegenden Fall wesentlich kleiner als die des Gruppen- bzw. Regionssimulationsmodells. Der auf der Basis des Einzelsimulationsmodells geschätzte Fehlerspielraum sollte allenfalls als *unterer Grenzwert* für den wahren Fehlerspielraum interpretiert werden. Das Ausmaß der stochastischen Variabilität hängt im Einzelsimulationsmodell stark von der Feinheit der Altersgliederung ab. Allgemeine Aussagen über die Ex-ante-Fehlervariabilität dieses Modells sind problematisch. Im Gegensatz dazu sind die Effekte der Altersstruktur im Regions- bzw. Gruppensimulationsmodell wesentlich kleiner. Aus inhaltlicher Sicht hat das Regions- bzw. Gruppenmodell den Vorteil, daß es regions- bzw. gruppenspezifische Verhaltensweisen berücksichtigt. In diesem Modell wird die altersspezifische Geburtenwahrscheinlichkeit jedes Individuums in einen kollektiven, d. h. regions- bzw. gruppentypischen Teil und in eine individuell bedingte personenabhängige Komponente zerlegt.

Die Berechnungen sollen auch für Städte bzw. Regionen anderer Einwohnergrößenklassen durchgeführt werden. Als Zwischenergebnis kann festgehalten werden, daß die Stadt- und Regionalplanung nicht mit einheitlichen Fehlerspielräumen für alle planungsrelevanten Altersgruppen rechnen darf. So liegt z. B. der Variationskoeffizient für die Zahl der Einwohner in der Altersgruppe über 65 Jahre (Krankenhausplanung, Altenhilfe usw.) bei einem Prognosezeitraum bis zu 30 Jahren zum Teil bei dem Dreifachen des Fehlerspielraums für die unter 15jährigen (Schulplanung).

Summary

(On stochastic variability of population projections for regions or population groups — comparison of alternative simulation models, using the town of Werne as a case study). Differences in population forecasts arise from three principal sources, the variability of the statistical information used, differences of assumption in the cause-and-effect relationships employed, and differences in treatment of the basically stochastic process of population change.

The stochastic effects occurring depend on the size of the region being considered, i. e. on the number of inhabitants. For the town of Werne of approximative 30,000 inhabitants two different model types (both without migration) have been investigated for their stochastic behaviour, the one a *micro simulation model*, and the other a *regional or group simulation model*. Whereas the expectation values of the forecasted or simulated population (according to age and gender) were essentially the same, the stochastic variability of the forecasts — as measured by the coefficient of variation obtained from a large number of simulation runs — was different. The stochastic variability of the result of the micro model was considerably less than that obtained from the group simulation model. The error range obtained from the micro model therefore has to be regarded as a *lower limit* for the actual variation that can occur.

The level of variability of the results of the micro model depends heavily on the degree of age distinction (age groups) used for the forecast. General statements on the variability are — *ex ante* — not possible for this model; the equivalent effects on the results of the group simulation type of model are much smaller.

The regional or group type of model has the decisive advantage that specific regional or group behaviour is taken into account; the age specific probabilities of births are separated into a collective (i. e. regional or group) component and an individual or personal component.

The research will be extended by performing similar calculations for towns and regions of differing population sizes. But the results achieved so far indicate that town and regional planning must reckon with the fact that population forecasts have differing levels of accuracy for the various age groups. For instance, the coefficient of variation for the forecasted number of old people (over 65) with relevance for hospital planning and aid for the old in same cases was about three times that for children (less than 15 years old) — relevance: school planning — for forecasts extending over up to 30 years.

1. Einleitung

Der Einsatz von leistungsfähigen Computern ermöglicht es, daß die Planungsträger von Regionen und Städten bis hin zu den kleinsten Gemeinden für ihre Bevölkerungsvorausschätzungen Prognosemodelle einsetzen. Aber die leistungsfähigen Rechner lassen den Benutzer im Stich, wenn es darum geht, die Berechnungsergebnisse zu interpretieren. So schließen beispielsweise Prognostiker aus dem Eintreffen von Vorausschätzungen, daß die getroffenen Prognoseannahmen richtig waren oder daß das Prognosemodell sich bewährt habe. Es wird dabei übersehen, daß mehrere Prognoseannahmen falsch sein können und daß das Prognoseergebnis gerade deshalb mit der Wirklichkeit übereinstimmen kann, weil sich mehrere falsche Annahmen kompensieren. Wird beispielsweise die Entwicklung der Geburtenhäufigkeit überschätzt (es

werden zu viele Geburten vorausgeschätzt) und die Entwicklung der Mortalität oder die Zahl der Wegzüge aus der Region unterschätzt, so kann die prognostizierte Bevölkerungszahl exakt mit der wirklichen übereinstimmen, obwohl — oder besser — weil die Annahmen falsch sind.

Eine Prognose ist identisch mit der Gesamtheit der in den inhaltlichen Prognoseannahmen enthaltenen Implikationen. Das Explizitmachen der in den Prognoseannahmen implizit enthaltenen Prognoseergebnisse ist bloße Technik, die den unwichtigsten Teil jeder Prognosearbeit ausmacht, auch wenn in der Praxis auf die Technik viel mehr Arbeitszeit und Mühe aufgewandt wird als auf das eigentliche Geschäft des Prognostikers, nämlich zutreffende Annahmen zu erarbeiten. Das Urteil über die Qualität einer Bevölkerungsprognose sollte daher nicht aus einem Vergleich der prognostizierten Bevölkerungszahl und -struktur mit der faktischen Entwicklung abgeleitet werden, sondern aus einem Vergleich der angenommenen Beziehungen zwischen den Variablen des Modells mit den tatsächlichen Zusammenhängen (bzw. Nicht-Zusammenhängen, was häufiger vorkommt) in der realen Welt.

Die Abweichungen zwischen den wirklichen und den prognostizierten Werten der Variablen beruhen allerdings nicht allein auf der Fehlerhaftigkeit der Annahmen über die Ursache-Wirkungsbeziehungen, sondern auch darauf, daß der Prozeß der realen Entwicklung stochastisch im Sinne von nicht-deterministisch ist. Die Aussage, daß der Prozeß der Realität stochastisch ist, kann weder bewiesen noch widerlegt werden, weil das Dingfestmachen von alternativen Zuständen der Welt, die außer dem realisierten Zustand auch möglich gewesen wären, naturgemäß daran scheitert, daß das nur Mögliche nicht real existiert. Das bedeutet nicht, daß es willkürlich ist, von der Prämisse einer stochastischen Welt auszugehen; es bedeutet aber, daß die Annahme einer stochastischen (i. S. von nicht-deterministischen) Realität den Rang einer Basisannahme hat. Diese Annahme wird heute einhellig akzeptiert, mit der Folge, daß sie als eine Selbstverständlichkeit gar nicht mehr genannt wird, wenn die Voraussetzungen eines Prognosemodells beschrieben werden. Für die vorliegende Untersuchung ist die Annahme der Stochastik wichtig, weil sich aus dem Möglichkeitscharakter der Welt eine eigenständige Quelle von Gründen für die Abweichungen zwischen Prognose und Realität ergibt.

Wir haben es bisher bewußt vermieden, den Begriff „Prognosefehler“ zu verwenden. Wenn eine Prognose trotz zutreffender Annahmen über Ursache-Wirkungsbeziehungen und trotz richtiger Ausgangsdaten nicht eintrifft, ist der Ausdruck „Fehler“ nicht zweckmäßig, denn dieser Begriff erweckt den Eindruck, daß sich die Gründe, die zu der Abweichung geführt haben, durch eine Beseitigung von Mängeln des Modells zumindest prinzipiell beseitigen lassen, während die Abweichung in Wahrheit auf einer fundamentalen, nicht veränderbaren Grundeigenschaft der Welt beruht. Statt des Begriffs „Fehler“ ist ein neutralerer Begriff vorzuziehen, z. B. „Diskrepanz“ oder „Divergenz“. Der Begriff „Fehler“ sollte zur Bezeichnung derjenigen Teilquellen der Diskrepanz zwischen Prognose und Realität verwendet werden, die wenigstens

im Prinzip vermeidbar sind, also auf Abweichungen, die auf fehlerhaften Daten und auf falschen Annahmen über die Ursache-Wirkungsbeziehungen beruhen. Für die Abweichungen, die auf der stochastischen Natur der Wirklichkeit beruhen, soll hier der Begriff „stochastische Variabilität“ verwendet werden. Wir unterscheiden also drei prinzipiell verschiedene Arten von Prognosefehlern: (1) Fehler, die auf falschen Analysedaten beruhen, (2) Fehler, die aus falschen Annahmen über Ursache-Wirkungsbeziehungen resultieren und (3) die Abweichungen zwischen Prognosemodell und Realität, die ihre Ursache in dem Möglichkeitscharakter der realen Entwicklung haben (vgl. Übersicht 1).

Übersicht 1
Überblick über die Fehlerarten bei Bevölkerungsprognosen

Ursache für die Diskrepanz zwischen Prognose und Realität	Bezeichnung des „Fehlers“
1. Unkorrekte Basisdaten 2. Unrichtige Annahmen über die Ursache-Wirkungsbeziehungen 3. Stochastische Variabilität des realen Prozesses	1. „Datenfehler“ 2. „Annahmenfehler“ = Prognosefehler i. e. S. 3. „Stochastische Varianz“
Ursachen insgesamt	„Prognosefehler i. w. S.“

Dies führt zu der Frage, wie sich die drei Fehlerkomponenten voneinander isolieren lassen. Da die drei Fehlerarten in den verschiedenen Bevölkerungsprognosemodellen nicht das gleiche Gewicht haben, muß die Frage für die verschiedenen Modelltypen gesondert untersucht werden. Folgende Gruppen von Modellen lassen sich unterscheiden:

(a) *Gliederung der Prognosemodelle nach der regionalen Dimension*

- regionale Modelle für eine einzelne Stadt oder Region,
- interregionale Modelle für mehrere Regionen, die durch demographische Verflechtungen (Wanderungen) oder durch wirtschaftliche Beziehungen (Handelsströme) miteinander verflochten sind.

(b) *Gliederung nach dem Umfang der abgebildeten Ursache-Wirkungsbeziehungen*

- rein demographische Modelle (ohne Einbeziehung von ökonomischen Variablen) mit oder ohne Wanderungen,
- demo-ökonomische Modelle mit Wirkungsbeziehungen zwischen demographischen und ökonomischen Variablen. Diese Modelle enthalten fast immer auch Wanderungen.

(c) *Gliederung nach der Technik der Ergebnisermittlung*

- deterministische Modelle, die Punktprognosen liefern,
- deterministische Modelle, die Intervallprognosen liefern, indem die Variabilität des Ergebnisses durch Setzung alternativer Prognoseannahmen (= Annahmensimulation) bzw. durch Ziehen von Zufallszahlen (Monte-Carlo-Verfahren) erreicht wird (Keyfitz, 1985),
- stochastische Modelle auf der Basis von Stichprobenpopulationen oder für die Totalpopulation einer Stadt bzw. einer Region, die Intervallprognosen liefern, wobei die stochastische Variabilität der Ergebnisse durch Ziehung von Zufallszahlen für einzelne Individuen erzeugt wird (= „Mikrosimulationsmodelle“).

Die unterschiedlichen Gliederungsgesichtspunkte schließen einander nicht aus. In den meisten Modellen sind die verschiedenen Gesichtspunkte miteinander kombiniert. Für kleinräumige Modelle ist der Gesichtspunkt des Umfangs der berücksichtigten Ursache-Wirkungsbeziehungen, insbesondere die Frage, ob Wanderungsströme enthalten sind, von entscheidender Bedeutung für die Ergebnisqualität. Aber auch dann, wenn kleinräumige Modelle Wanderungsströme enthalten, variiert die Qualität der Modelle stark in Abhängigkeit vom Realitätsgehalt der abgebildeten Zusammenhänge zwischen den demographischen und den ökonomischen Variablen. Entscheidend für die Qualität der regionalen und interregionalen Modelle ist die Art, wie die Rückkopplungen zwischen den verschiedenen Komponenten der regionalen Arbeitsmarktbilanz abgebildet sind. Denn die Arbeitsmarktbilanz enthält sämtliche demographisch relevanten Variablen sowie viele der sozio-ökonomischen Größen, von denen die demographischen Variablen abhängen. Beispielsweise ist das Fertilitätsniveau einer Region (u. a.) von den Zu- und Wegzügen abhängig, weil über familiäre Veränderungen und berufliche Veränderungen bei Wohnortwechseln häufig eine simultane Entscheidung getroffen wird. Die Arbeitsmarktbilanz enthält folgende Variablen, die in einem regionalen Bevölkerungsprognosemodell berücksichtigt werden sollten (Birg, 1979):

I. *Variablen, die das Angebot an Arbeit beeinflussen:*

- (1) Veränderung des Erwerbspotentials in der Region auf Grund der natürlichen Bevölkerungsentwicklung durch Geburten und Sterbefälle zuzüglich der entsprechenden Erwerbsquote,
- (2) zugezogene Personen aus anderen Regionen bzw. aus dem Ausland zuzüglich der Erwerbsquoten der entsprechenden Teilbevölkerungen,
- (3) fortgezogene Personen nach anderen Regionen bzw. in das Ausland zuzüglich der Erwerbsquoten der entsprechenden Teilbevölkerungen,
- (4) Zahl der Einpendler und Auspendler.

II. Variablen, die die Nachfrage nach Arbeit beeinflussen:

- (5) Zahl der zusätzlichen Arbeitsplätze durch Erweiterung bestehender Betriebe oder durch Zuzug oder Gründung neuer Betriebe,
- (6) Zahl der Arbeitsplätze, die durch Rationalisierungsmaßnahmen oder durch Betriebsschließungen und den Wegzug von Betrieben verloren gingen.

III. Variablen zur Beschreibung des Saldos der regionalen Arbeitsmarktbilanz:

- (7) Veränderung der „stillen Reserve“ des Arbeitskräfteangebots,
- (8) Veränderung der Zahl der Arbeitslosen,
- (9) Veränderung der Zahl der offenen Stellen.

Unter diesen 9 Variablengruppen sind die Zuzüge und die Fortzüge einschließlich der Erwerbsquoten der entsprechenden Bevölkerungsströme die weitaus wichtigsten Komponenten der Bevölkerungsbilanz bzw. der Arbeitsmarktbilanz: Die Zahl der Zuzüge ist bei kreisfreien Städten und Landkreisen um ein Vielfaches größer als die Zahl der Geburten, die Zahl der Fortzüge um ein Vielfaches größer als die Sterbefälle (Birg, 1979, S. 91). Je kleiner die Regionen bzw. Städte sind, desto größer ist das Verhältnis der Zuzüge zu den Geburten bzw. der Fortzüge zu den Sterbefällen. Die Annahmensetzungen bezüglich der Zu- und Fortzüge bergen daher die größten Fehlerquellen in sich. Im Vergleich zu diesen Annahmefehlern ist das Ausmaß von Fehlern in bezug auf die unterstellte Entwicklung der Mortalität und Fertilität gering.

Es macht wenig Sinn, den Fehlerspielraum von Prognosemodellen zu vergleichen, die eine voneinander abweichende Variablen- bzw. Annahmenstruktur haben. Im folgenden sollen die Fehlerspielräume von zwei Simulations- bzw. Prognosemodellen für den einfachsten Fall miteinander verglichen werden, nämlich für einen Modelltyp, der lediglich die natürliche Bevölkerungsentwicklung abbildet (ohne Wanderungen). Die Arbeiten sind als ein erster Schritt zu werten. Die wichtigsten Variablen, nämlich die Wanderungen, sollen in weiteren Berechnungen einbezogen werden. Es ist geplant, die Untersuchungen über das Fallbeispiel (Stadt Werne) hinaus auf Städte bzw. Regionen unterschiedlicher Größe auszuweiten.

2. Grundlagen stochastischer Bevölkerungsprognosemodelle

(a) Das deterministische Leslie-Modell

Grundlage der in der Praxis verwendeten Vorausschätzungsmodelle sind drei Klassen von Informationen:

1. Daten für den nach Alter und Geschlecht differenzierten Bevölkerungsbestand der Frauen und der Männer im Basisjahr der Prognose:

$F_{a,t}$ bzw. $M_{a,t}$ Zahl der Frauen (F) bzw. der Männer (M) im Alter a im Basisjahr t .

2. Die nach Alter und Geschlecht differenzierten Sterbewahrscheinlichkeiten und Überlebenswahrscheinlichkeiten:

$q_{a,t}^w, q_{a,t}^m$ Wahrscheinlichkeit für eine Frau (w) bzw. für einen Mann (m), die exakt a Jahre alt sind, vor Erreichen des Alters $a+1$ zu sterben.

$p_{a,t}^w = 1 - q_{a,t}^w$ Überlebenswahrscheinlichkeit der Frauen.

$p_{a,t}^m = 1 - q_{a,t}^m$ Überlebenswahrscheinlichkeit der Männer.

3. Die nach einzelnen Altersjahren differenzierten Geburtenziffern

$b_{a,t} = \frac{G}{F_{a,t}} 1000$ Zahl der Geburten (G) auf 1000 Frauen im Alter a .

$b_{a,t}^w = \frac{G^w}{F_{a,t}} 1000$ Zahl der Mädchengeburt (G^w) auf 1000 Frauen im Alter a .

Die Sterbewahrscheinlichkeiten, die der Sterbetafel entnommen werden, sind für Personen im exakten Alter a definiert, also auch nur anwendbar auf Personen, die genau 0, 1, 2, ... Jahre alt sind. Da der Bevölkerungsbestand realer Bevölkerungen in Altersintervalle von 0 bis unter 1, 1 bis unter 2 usw. gliedert ist, können die Sterbewahrscheinlichkeiten erst nach Umrechnung in Sterbeziffern für die Prognose verwendet werden. Als Sterbeziffer für Personen im Altersintervall von a bis unter $a + 1$ Jahren wird hier das arithmetische Mittel aus den Sterbewahrscheinlichkeiten q_a und q_{a+1} definiert (mit Ausnahme des Altersintervalls 0 bis u. 1, bei dem der Verlauf stark konkav ist). Im folgenden dient das Symbol q sowohl für die Kennzeichnung der Sterbewahrscheinlichkeit als auch für die Sterbeziffer; die jeweilige Bedeutung ergibt sich aus dem Zusammenhang. Bei den Geburten ist im Prinzip eine analoge begriffliche Unterscheidung zwischen Geburtenwahrscheinlichkeit und Geburtenziffer erforderlich. Die Statistischen Ämter berechnen und publizieren jedoch nur Geburtenziffern. Sie stellen keine „Geburtenwahrscheinlichkeiten“ auf, weil sich das generative Verhalten im Gegensatz zur Mortalität rasch und intensiv ändern kann, wie der jüngste Ge-

burtenrückgang (1965—75) gezeigt hat, bei dem sich die Geburtenzahlen in 10 Jahren fast halbierten.

Die einzelnen Rechenschritte zur Fortschreibung des weiblichen Bevölkerungsbestandes vom Basisjahr t zum Basisjahr $t + 1$ lassen sich am einfachsten durch folgende Matrix-Schreibweise darstellen:

$$(1) \begin{bmatrix} F_{0,t+1} \\ F_{1,t+1} \\ \cdot \\ \cdot \\ F_{m,t+1} \end{bmatrix} = \begin{bmatrix} b_{0,t}^w & b_{1,t}^w & \cdot & \cdot & \cdot & b_{m,t}^w \\ p_{0,t}^w & & & & & \\ & p_{1,t}^w & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & p_{m-1,t}^w & 0 \end{bmatrix} \begin{bmatrix} F_{0,t} \\ F_{1,t} \\ \cdot \\ \cdot \\ F_{m,t} \end{bmatrix}$$

bzw.

$$(2) F_{t+1} = L_t^w F_t.$$

Die Matrix L_t^w , die nach P. H. Leslie auch als Leslie-Matrix oder als Projektionsmatrix bezeichnet wird, enthält in der ersten Zeile die altersspezifischen Geburtenziffern $b_{a,t}^w$, die die Zahl der Mädchengeburten auf 1000 Frauen angeben, wobei die Elemente für das Alter vor 14 und für das Alter nach der Reproduktionsphase (45) Null sind. In der Sub-Diagonalen stehen die Überlebenswahrscheinlichkeiten. Alle übrigen Elemente der Matrix sind Null. Durch die Anordnung der Überlebenswahrscheinlichkeiten in der Sub-Diagonalen wird bei der Multiplikation der einzelnen Matrixzeilen mit dem Bestandsvektor F_t nicht nur die Zahl der Überlebenden berechnet, sondern gleichzeitig der Altersprozeß abgebildet, indem der Altersindex automatisch um eine Einheit erhöht wird.

Das erste Element des Vektors F_{t+1} gibt die Zahl der Mädchengeburten an. Durch Multiplikation dieser Zahl mit der Relation von Knaben- zu Mädchengeburten (beispielsweise 105:100) erhält man das erste Element im Vektor M_{t+1} für den männlichen Bevölkerungsbestand. Die zeitliche Entwicklung der Bestandsvektoren M_t läßt sich mit einer zu Gleichung (2) analogen Matrix-Gleichung beschreiben:

$$(3) M_{t+1} = L_t^m M_t.$$

Die Leslie-Matrix L_t^m für die Männer enthält analog zur Matrix L_t^w für die Frauen in der Sub-Diagonalen die Überlebenswahrscheinlichkeit für die Männer. Alle übrigen Elemente sind Null.

Sind die Geburten- und Sterbeziffern im Prognosezeitraum konstant, so reduziert sich die Bevölkerungsprognose auf eine Fortschreibung des Basisbestandes durch wiederholte Anwendung der Rekursionsgleichungen (2) und (3) für aufeinanderfolgende Jahre. Nach n -maliger Fortschreibung erhält man den Bevölkerungsbestand im Jahr $t + n$:

$$(4) \quad \begin{aligned} F_{t+n} &= (L^w)^n F_t; \\ M_{t+n} &= (L^m)^n M_t. \end{aligned}$$

In diesen Modellen ist die Zahl der Geburten eine Funktion der Zahl und Altersstruktur der Frauen und der Geburtenziffern, nicht eine Funktion der Zahl der Männer (= „eingeschlechtliche“ Modelle). Nur selten wird bei der Vorausschätzung der Geburtenzahl auch die Zahl der Männer berücksichtigt (= „zweigeschlechtliche“ Modelle), was auf die Schwierigkeiten zurückzuführen ist, diese Modelle analytisch zu handhaben. In numerischen Mikrosimulationsmodellen, wie sie im folgenden verwendet werden, bereitet es keine Probleme, zweigeschlechtliche Ansätze zu formulieren, doch lohnt sich der damit verbundene Aufwand im Hinblick auf die hier interessierenden Fragen nicht.

(b) Stochastische Leslie-Modelle

Das deterministische Leslie-Modell kann auf verschiedene Weise stochastisiert werden. Ziel der Stochastisierung ist die Ableitung der Erwartungswerte und der Varianzen der Prognoseergebnisse. Für die analytische Ableitung wird das Leslie-Modell meist auf die vorreproduktiven und auf die reproduktiven Altersgruppen der Frauen eingeschränkt, weil die Zahl der Frauen, die das reproduktionsfähige Alter überschritten hat, die künftige Bevölkerungszahl nicht durch Geburten beeinflusst.

Bezeichnet man die für die auf die Alter 0 bis 45 reduzierte Leslie-Matrix der Frauen mit A und die auf diese Altersgruppe reduzierten Bestandsvektoren mit x_t , so lassen sich die in der Literatur diskutierten Möglichkeiten, das Leslie-Modell zu stochastisieren, wie folgt darstellen.

(b₁) Stochastisierung durch Addition eines Zufallsterms u :

$$(5) \quad x_{t+1} = Ax_t + u_t.$$

Diese Form der Stochastisierung wurde von Sykes (1969) gewählt. Für die Berechnung der Konfidenzintervalle bzw. der Variationskoeffizienten der Elemente von x_{t+1} nimmt Sykes an, daß die Elemente der Zufallsvektoren u_t einen Erwartungswert von Null und eine konstante Varianz-Kovarianz-Matrix haben:

$$(6) \quad E(u_t) = 0$$

$$(7) \quad E(u_s u_t) = \Gamma_{st} = \Gamma, \quad s, t = 0, 1, \dots$$

Sykes hat die Varianz-Kovarianz-Matrix geschätzt, indem er für einen Zeitraum in der Vergangenheit, für den die Bevölkerungsentwicklung bekannt ist, eine Ex-post-Prognose für 15 aufeinanderfolgende Jahre durchführte. Daraus bildete er eine Sequenz von 15 Prognoseabweichungen, auf deren Basis er die Varianz-Kovarianz-Matrix ermittelte.

Dieser Ansatz ist für die Analyse der hier interessierenden Fehlerart, die wir als die stochastische Variabilität von Bevölkerungsprognosen bezeichnet haben (vgl. Übersicht 1), nicht geeignet, denn in den Differenzen zwischen den Ex-post-Prognosewerten und den realisierten Werten schlagen sich sämtliche Fehlerursachen nieder, insbesondere die Annahmefehler, die die größte Bedeutung haben. Annahmefehler liegen vor, wenn — wie von Sykes vorausgesetzt — die Geburten- und Sterbeziffern als konstant angenommen werden (konstante Leslie-Matrix A), denn die Konstanz-Annahme bezüglich der Leslie-Matrix ist auf zweifache Weise unzutreffend. Die Geburten- und Sterbeziffern in der Leslie-Matrix sind eine Funktion der Geburten- und Sterbeziffern zweier verschiedener Bevölkerungen, erstens der Bevölkerung, die schon zu Beginn des Prognosezeitraums in dem betrachteten Gebiet lebte (bzw. bei nicht regionalen Prognosen: die Bevölkerung, die schon zu Beginn des Prognosezeitraums der betrachteten Bevölkerungsgruppe angehörte), zweitens der Bevölkerung, die später zugezogen ist (bzw. zur Gruppe dazugekommen). Die Leslie-Matrix ändert sich, wenn sich das Verhalten (Geburten- und Sterbeziffern) der ansässigen Bevölkerung im Prognosezeitraum ändert. Aber selbst dann, wenn sich das Verhalten der ansässigen Bevölkerung nicht ändert, ist die Leslie-Matrix in der Regel nicht konstant, wie folgende Überlegung zeigt: Wenn die Geburten- und Sterbeziffern der Migranten mit denen der ansässigen Bevölkerung nicht übereinstimmen, sind die Geburten- und Sterbeziffern in der Leslie-Matrix eine Funktion der Wanderungsströme. Da Wanderungen in regionalen Modellen eine große Bedeutung haben, ist die Annahme einer konstanten Leslie-Matrix tendenziell um so weniger erfüllt, je kleiner das betrachtete Gebiet ist.

(b₂) Bei dem folgenden Modelltyp wird die Stochastisierung des Leslie-Modells nicht durch Addition eines Zufallsvektors u_t zum Bevölkerungsvektor x_t erreicht, sondern durch Addition einer Zufallsmatrix Δ_t zur Leslie-Matrix:

$$(8) \quad x_{t+1} = (A + \Delta_t)x_t.$$

Der Ansatz wird in der Literatur unzweckmäßigerweise als Ansatz auf der Basis einer „zufälligen Leslie-Matrix“ bezeichnet, obwohl die Leslie-Matrix A hier ebenso wie in dem in Gleichung (5) dargestellten Modell als konstant angenommen wird. Sykes hat sowohl auf der Basis des in Gleichung (5) als auch auf der Basis des in Gleichung (8) dargestellten Modells Variationskoeffizienten für die Prognosefehler berechnet. Die Kritik, daß die Variationskoeffizienten auch die Annahmefehler enthalten, die aus der Konstanz-Annahme der Leslie-Matrix resultieren, trifft auch auf den in Gleichung (8) dargestellten Ansatz von Sykes zu.

(c) Verzweigungsprozeßmodelle

Die Modelle des Leslie-Typs werden in der Praxis in der Regel auf Gruppen von Personen angewandt. Daneben sind Modelle im Einsatz, bei denen die Be-

rechnungen für jedes Individuum getrennt vorgenommen werden. Zur Unterscheidung wollen wir hier von *Gruppenmodellen* und von *Einzelmodellen* sprechen.

Die Ergebnisse von Einzelmodellen werden entweder analytisch abgeleitet oder durch numerische Simulationsmethoden herausgearbeitet. Für einen bestimmten Typ der Einzelmodelle, die „*Verzweigungsprozeßmodelle*“, konnten von Sykes, Pollard und Feichtinger wichtige Eigenschaften analytisch abgeleitet werden. Grundlage der Verzweigungsprozeßmodelle ist es, die Abfolge der einzelnen Geburten bzw. Sterbefälle in der Zeit als einen Bernoulli-Prozeß zu interpretieren. Ein Bernoulli-Prozeß entsteht, wenn ein Experiment, bei dem nur zwei verschiedene, sich gegenseitig ausschließende Ereignisse eintreffen können, mehrmals wiederholt wird (z. B. Münzwürfe etc.). Die Analogie zum Geburtenprozeß besteht darin, daß das Ereignispaar „Geburt im Jahr t“ und „keine Geburt im Jahr t“ als die möglichen Realisationen eines Experiments interpretiert werden, das n-mal wiederholt wird, wobei n die Zahl der Frauen angibt, deren Geburtenzahl geschätzt werden soll. Sind in der betrachteten Gruppe n Frauen und hat jede Frau die gleiche Geburtenwahrscheinlichkeit f, dann ist die Zahl der Geburten eine binomialverteilte Zufallsvariable, vorausgesetzt, daß die individuellen Ereignisse stochastisch unabhängig sind. Nach der gleichen Analogie läßt sich auch der Sterbeprozess als ein Bernoulli-Prozeß mit einer Abfolge von Experimenten interpretieren, von denen jedes zwei sich gegenseitig ausschließende Ergebnisse (Tod bzw. Überleben) hat, wobei wiederum stochastische Unabhängigkeit vorausgesetzt werden muß. Die Zahl der Geburten G und die Zahl der Sterbefälle S haben unter dieser Voraussetzung folgende Erwartungswerte (E) und Varianzen (V):

$$(9) \quad E(G) = nf; \quad E(S) = nq;$$

$$(10) \quad V(G) = nf(1-f); \quad V(S) = nq(1-q).$$

Hier sind die *Variationskoeffizienten* (VK) der Variablen G und S von besonderem Interesse, weil wir die stochastische Varianz der Prognose in dieser Untersuchung durch den Variationskoeffizienten messen. Die Variationskoeffizienten der Zahl der Geburten VK_G bzw. der Zahl der Sterbefälle VK_S sind:

$$(11) \quad VK_G = \frac{\sigma_G}{\mu_G} = \sqrt{\frac{(1-f)}{nf}};$$

$$(12) \quad VK_S = \frac{\sigma_S}{\mu_S} = \sqrt{\frac{(1-q)}{nq}}.$$

Die Aufgabe, aus den Binomialverteilungen für den Geburten- und den Sterbeprozess und aus dem Vektor des Bevölkerungsbestandes für die Frauen die Erwartungswerte und die Varianzen der Elemente des Bestandsvektors am Ende des Prognosezeitraums zu berechnen, entspricht der hier vorliegenden

Fragestellung wesentlich besser als der oben dargestellte Ansatz der Ex-post-Varianzschätzung des Prognosefehlers von Sykes, denn für die Verzweigungsprozeßmodelle läßt sich eine Ex-ante-Fehleranalyse durchführen, ohne daß auf die Differenzen zwischen Modell und Realität in der Vergangenheit zurückgegriffen werden muß.

Der Ansatz hat allerdings einen Nachteil, der seine Eignung für die Untersuchung der Frage, wie die Regionsgröße die stochastische Variabilität der Prognose beeinflusst, einschränkt; denn um eine analytische Aussage treffen zu können, muß erstens für alle Individuen die gleiche Geburten- bzw. Sterbewahrscheinlichkeit unterstellt werden, und es muß zweitens angenommen werden, daß die individuellen Ereignisse stochastisch unabhängig voneinander sind — eine Voraussetzung, die um so weniger erfüllt ist, je stärker die Argumente sind, die dazu berechtigen, von regions- bzw. gruppenspezifischen Verhaltensweisen zu sprechen; denn wenn die Menschen in einer Region ähnliche Verhaltensweisen haben, dann beruht die Ähnlichkeit in der Regel gerade darauf, daß die Menschen sich *nicht* stochastisch unabhängig voneinander verhalten, sondern direkt oder indirekt miteinander in Beziehung stehen.

Sykes (1969), Pollard (1966) und Feichtinger (1971) haben gezeigt, wie sich unter der Annahme konstanter und stochastisch unabhängiger Ereignisse der Erwartungswert und die Standardabweichung bzw. der Variationskoeffizient als Quotient aus Standardabweichung und Erwartungswert analytisch ableiten lassen. Da der Variationskoeffizient für die Geburten (Gleichung 11) bzw. für die Sterbefälle (Gleichung 12) mit wachsendem n sinkt, ist es nicht überraschend, daß in den empirischen Anwendungen Variationskoeffizienten für die prognostizierte Bevölkerungszahl errechnet wurden, die um den Faktor 1000 kleiner sind als die Variationskoeffizienten, die sich aus der Ex-post-Varianzschätzung auf der Basis des Leslie-Modells ergeben. Für Bevölkerungen mit mehreren Millionen Frauen führt die Abfolge von Bernoulli-Experimenten zu einer extrem kleinen stochastischen Variabilität des Prognosefehlers. Feichtinger zieht daraus in Übereinstimmung mit Sykes den Schluß „...“, daß die beobachtete Variabilität bei realen Bevölkerungsentwicklungen nur zu einem *geringen* Teil durch die Unterstellung unabhängiger Binomialversuche erklärt werden kann“ (Feichtinger, 1971, S. 348), was zweifellos für große Bevölkerungen zutrifft. Aber wie verhält es sich mit kleinen Bevölkerungen für Städte und Regionen?

Auf diese Frage soll hier anhand des Fallbeispiels der Stadt Werne eine Antwort versucht werden. Dabei werden zwei verschiedene Prognose- bzw. Simulationsmodelle im Hinblick auf die stochastische Variabilität ihrer Ergebnisse miteinander verglichen, und zwar ein *Einzelsimulationsmodell*, das auf der Annahme stochastisch unabhängiger Individualereignisse beruht, und ein *Regions- bzw. Gruppensimulationsmodell*, das diese restriktive Annahme nicht enthält.

3. Ableitung eines stochastischen Modells zur Prognose und Simulation der Bevölkerungsentwicklung von Regionen bzw. Bevölkerungsgruppen ohne Wanderungen

Wir nehmen an, daß die Individuen in der Bevölkerung einer Region oder allgemeiner in einer Bevölkerungsgruppe einerseits Verhaltensähnlichkeiten haben, die allen Individuen gemeinsam sind, sich aber andererseits prinzipiell in ihrem Verhalten voneinander unterscheiden. Um beide Gesichtspunkte zu berücksichtigen, setzen wir voraus, daß sich die individuelle Wahrscheinlichkeit dafür, daß sich ein Individuum in einer bestimmten Weise verhält, bzw. daß ihm ein bestimmtes Ereignis widerfährt, aus zwei Komponenten zusammensetzt, einem regionsspezifischen (bzw. gruppenspezifischen) Teil, den wir auch als kollektive Verhaltenskomponente bezeichnen, und einem individuellen Teil. Wenden wir diese Voraussetzung auf die Geburtenwahrscheinlichkeit an, die wir zur Elimination des Effektes der Altersstruktur von vornherein nach einzelnen Altersjahren a differenzieren, so setzt sich die Wahrscheinlichkeit für das Individuum i , im Jahr t ein Kind zur Welt zu bringen, aus einem kollektiven und aus einem individuellen Term zusammen:

$$(13) \quad f_{t,a}^i = f_{0,a}^w + f_{0,a}^w y_{t,a}^i; \quad i = 1, 2, \dots, F_{t,a}.$$

Hier gibt $f_{0,a}^w$ den Teil der Geburtenwahrscheinlichkeit an, der allen Individuen aus der Gruppe w (hier: allen Frauen der Stadt Werne) gemeinsam ist (= kollektive Komponente), während der zweite Term den Teil der Geburtenwahrscheinlichkeit spezifiziert, der für alle Individuen verschieden ist (individuelle Komponente). Der individuelle Teil wird als Differenz zwischen der individuellen Geburtenwahrscheinlichkeit und der kollektiven Komponente der Geburtenwahrscheinlichkeit definiert, wobei die Differenz durch Multiplikation des kollektiven Teils mit einem individuell verschiedenen Faktor $y_{t,a}^i$ (= „Individualfaktor“) bestimmt wird:

$$f_{0,a}^w y_{t,a}^i.$$

Für den Individualfaktor nehmen wir an, daß er eine stochastische Größe ist, die einer individuellen Verteilung folgt. Ist $y_{t,a}^i = 0,20$, so bedeutet das, daß die Geburtenwahrscheinlichkeit des betreffenden Individuums im Jahr t um 20 % über dem kollektiven Wert liegt. Die Individualfaktoren sollen stochastisch unabhängig voneinander sein. Über die Verteilungsform der $y_{t,a}^i$ treffen wir keine Annahmen. Wir nehmen allerdings an, daß die Erwartungswerte der Individualfaktoren sich voneinander unterscheiden. Ist der Erwartungswert des Individualfaktors für ein Individuum größer als Null, so ist die Geburtenwahrscheinlichkeit größer als die kollektive Komponente $f_{0,a}^w$ und umgekehrt. In dem von Null verschiedenen Erwartungswert des Individualfaktors schlagen sich die individuellen Besonderheiten der Person nieder, soweit sie für das generative Verhalten von Bedeutung sind. Die individuell verschiedenen Erwartungswerte der Individualfaktoren seien:

$$(14) \quad E(y_{t,a}^1) + E(y_{t,a}^2) + \dots + E(y_{t,a}^i).$$

Obwohl wir individuell verschiedene bzw. von Null abweichende Erwartungswerte der Individualfaktoren voraussetzen, nehmen wir gleichzeitig an, daß der Erwartungswert der *Summe* der Individualfaktoren Null ist. Mit dieser wichtigen Annahme erreichen wir eine Normierung unserer Aussagen über die Höhe der regionsgrößenbedingten Variabilität der regionalen Bevölkerungsprognosen, die in dieser Arbeit im Zentrum des Interesses steht. Wenn nämlich

$$(15) \quad E(\sum_i y_{t,a}^i) = 0 \quad \text{für alle } a \text{ und } t$$

vorausgesetzt wird, dann ist die regionsgrößenbedingte Variabilität der Prognosen um so kleiner, je größer die Region ist, und wenn wir die betrachtete Region immer größer werden lassen, bis sie mit dem Gesamttraum identisch ist, dann ist die regionsgrößenbedingte Variabilität des Prognoseergebnisses definitionsgemäß Null, wie sich unten zeigen wird. Diese sinnvolle Implikation der Annahmen für die Eichung der Variabilität sei hier festgehalten, weil sie für das Verständnis der weiteren Ableitung wichtig ist.

Wir wollen nun auf der Basis der getroffenen Annahmen einen Schätzwert für den Erwartungswert und für die Varianz der regionalen altersspezifischen Geburtenziffern ableiten. Die regionale Varianz einer altersspezifischen Geburtenziffer ist wichtig, weil wir aus ihr Aussagen über die regionsgrößenbedingte Variabilität der Bevölkerungsprognosen ermitteln. Der Erwartungswert der altersspezifischen Geburtenziffer der Region w als Ganzes ist der Erwartungswert des Quotienten aus der Zahl der Geburten in der Region und der Zahl der Frauen in der Region:

$$(16) \quad E(f_{t,a}^w) = E\left[\frac{G_{t,a}^w}{F_{t,a}^w}\right] = \frac{E(G_{t,a}^w)}{F_{t,a}^w}.$$

Hierbei ist $F_{t,a}^w$ die Zahl der Frauen des Alters a im Jahr t in der Region w und $E(G_{t,a}^w)$ der Erwartungswert der Zahl der Geburten. Die zu erwartende Geburtenzahl ist die Summe der individuellen Geburtenwahrscheinlichkeiten bzw. mit (13)

$$\begin{aligned} (17) \quad E(G_{t,a}^w) &= E(\sum_i f_{t,a}^i) = E[\sum_i (f_{0,a}^w + f_{0,a} y_{t,a}^i)] \\ &= E(\sum_i f_{0,a}^w) + E(\sum_i f_{0,a}^w y_{t,a}^i) \\ &= E(F_{t,a}^w f_{0,a}^w) + f_{0,a}^w E(\sum_i y_{t,a}^i) \\ &= F_{t,a}^w f_{0,a}^w + f_{0,a}^w E(\sum_i y_{t,a}^i). \end{aligned}$$

Substituieren wir (17) in (16), so ergibt sich, daß der Erwartungswert der regressionspezifischen Geburtenziffer mit dem kollektiven Teil der individuellen Geburtenziffer identisch ist:

$$(16.1) \quad E(f_{t,a}^w) = f_{0,a}^w + f_{0,a}^w \frac{E(\sum_i y_{t,a}^i)}{F_{t,a}^w} \\ = f_{0,a}^w$$

Zur Abkürzung der Schreibweise definieren wir

$$(18) \quad z_{t,a} = \frac{\sum_i y_{t,a}^i}{F_{t,a}^w}, \text{ wobei}$$

$$E(z_{t,a}) = 0;$$

$$V(z_{t,a}) = \sigma_{t,a}^2.$$

Der Erwartungswert der stochastischen Größe $z_{t,a}$ ist wegen (15) in Verbindung mit der Annahme stochastisch unabhängiger Individualfaktoren Null:

$$(19) \quad E(\sum_i y_{t,a}^i) = \sum_i (E(y_{t,a}^i)) = 0.$$

Die Varianz der Geburtenziffer der Region ist mit Definition (18)

$$(20) \quad V(f_{t,a}^w) = V(f_{0,a}^w + f_{0,a}^w z_{t,a}) = (f_{0,a}^w)^2 \sigma_{t,a}^2.$$

Nehmen wir der Einfachheit halber an, daß $V(z_{t,a})$ für alle Altersjahre gleich ist:

$$(21) \quad \sigma_{t,a}^2 = \sigma_t^2,$$

erhalten wir

$$(22) \quad V(f_{t,a}^w) = E(f_{t,a}^w - f_{0,a}^w)^2 = (f_{0,a}^w)^2 \sigma_t^2.$$

Summation von Gleichung (22) über alle Alter ergibt

$$(23) \quad \sum_a [E(f_{t,a}^w - f_{0,a}^w)^2] = \sigma_t^2 \sum_a (f_{0,a}^w)^2$$

bzw.

$$(24) \quad \sigma_t^2 = \frac{\sum_a [E(f_{t,a}^w - f_{0,a}^w)^2]}{\sum_a (f_{0,a}^w)^2} = \frac{E[\sum_a (f_{t,a}^w - f_{0,a}^w)^2]}{\sum_a (f_{0,a}^w)^2}.$$

Somit ist $\hat{\sigma}_t^2$ ein Schätzwert für die unbekannte Varianz σ_t^2 :

$$(25) \quad \hat{\sigma}_t^2 = \frac{\sum_a (f_{t,a}^w - f_{0,a}^w)^2}{\sum_a (f_{0,a}^w)^2}.$$

Um $\hat{\sigma}_t^2$ empirisch ermitteln zu können, benötigen wir neben den für die Region gemessenen Geburtenziffern $f_{t,a}^w$ die unbekanntenen „wahren“ altersspezifischen Geburtenziffern $f_{0,a}^w$ für die Regionsbevölkerung, die mit dem kollektiven Teil der individuellen Geburtenwahrscheinlichkeiten identisch sind. Für den Unterschied zwischen den „wahren“ bzw. kollektiven Ziffern und den beobachteten Ziffern sind die stochastischen Komponenten der individuellen Ge-

burtenziffern verantwortlich. Da der Erwartungswert der in den verschiedenen Jahren beobachteten Geburtenziffern $f_{t,a}^w, f_{t+1,a}^w, \dots, f_{t+n,a}^w$ mit der „wahren“ Geburtenziffer $f_{0,a}^w$ identisch ist, können wir den aus mehreren Jahren gemittelten Durchschnitt der Ziffern als Schätzwert für $f_{0,a}^w$ verwenden, vorausgesetzt, daß Meßwerte für mehrere Jahre vorliegen. Im vorliegenden Fall, d. h. für die Stadt Werne, waren Geburtenziffern nur für das Jahr 1981 verfügbar. Wir haben deshalb anstelle der aus mehreren Jahren gemittelten Geburtenziffern die geglätteten Werte der altersspezifischen Geburtenziffern für das Jahr 1981 verwendet, wobei als approximierende Funktion die Coale-Trussel-Funktion zugrunde gelegt wurde (Coale u. Trussel, 1974 u. 1978). Alternativ wurden auch die Beta-Verteilung und die Gamma-Verteilung als approximierende Funktionen geprüft, doch war die Anpassung der Coale-Trussel-Funktion am besten (vgl. Schaubilder 1 u. 2). Für die Varianz der Geburtenziffern entsprechend Gleichung (25) ergaben sich folgende Werte:

$$\sum_a f_{81,a}^2 \quad \hat{\sigma}_{81} = \sqrt{\hat{\sigma}_{81}^2}$$

Beta-Verteilung	148 186	0,18848
Gamma-Verteilung	143 070	0,21600
Coale-Trussel-Funktion	141 215	0,19250

Das Geburtenziffernmuster für die Bundesrepublik insgesamt ist als approximierende Funktion nicht geeignet, weil die Geburtenziffern in Werne bis zum Alter 29 sämtlich über den Werten für die Bundesrepublik liegen (Vergleich auf der Basis der Dichtefunktionswerte für beide Muster), für die Alter über 29 darunter.

Schaubild 1:
Geburtenziffern Werne und Bundesrepublik 1981

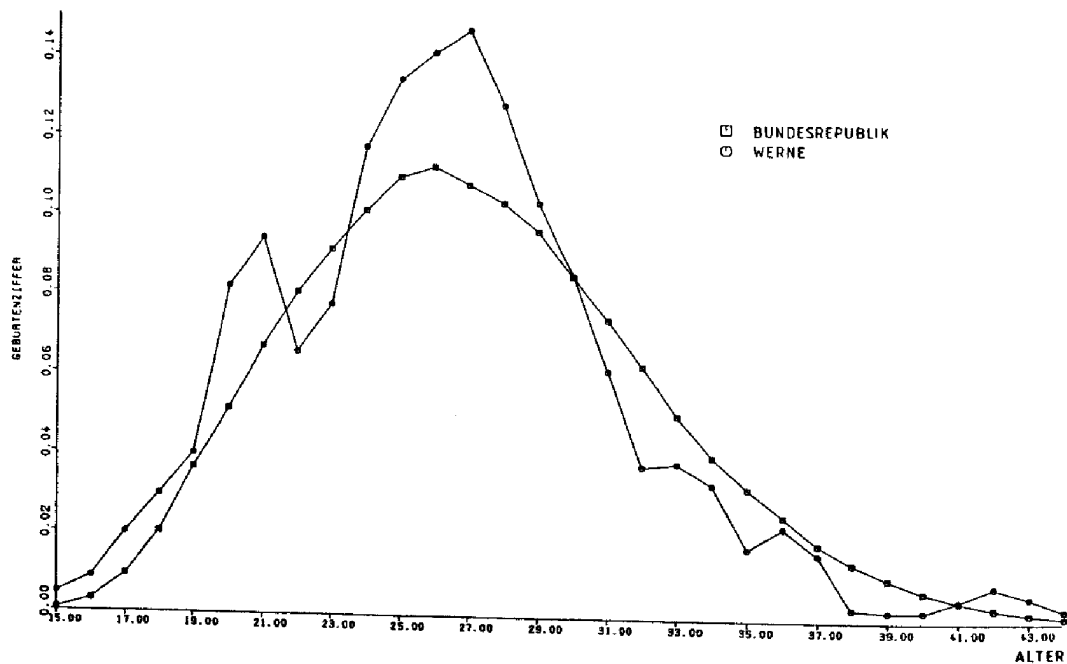
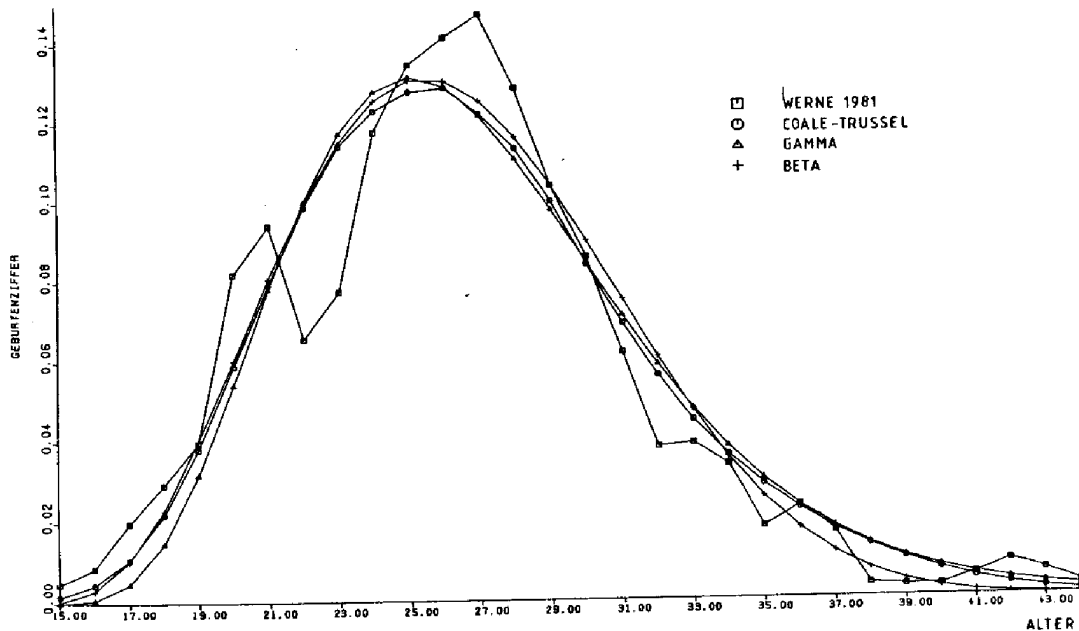


Schaubild 2:
Geburtenziffern Werne 1981 und approximierende Funktionen

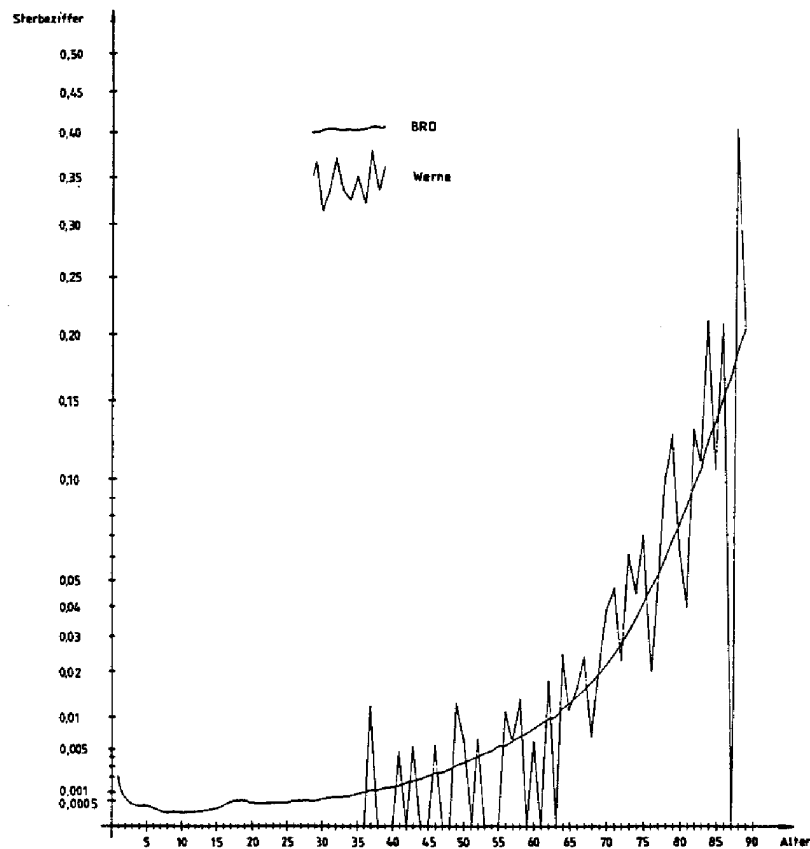


4. Ergebnisse der Simulationsstudie für die Stadt Werne

Die stochastische Variabilität einer regionalen Bevölkerungsprognose hängt nicht nur von den Eigenschaften des betrachteten Gebiets ab, beispielsweise von der absoluten Einwohnerzahl und vom Verhalten der Einwohner, sondern auch von den Eigenschaften des Prognosemodells. Obwohl hier primär der Einfluß der Regionsgröße interessiert, wurde auch der Frage nachgegangen, ob es typische Effekte der Modellstruktur gibt — eine Vermutung, die sich bestätigt hat. Die dem Modellvergleich zugrunde gelegten Modelle sind das *Einzel-simulationsmodell* (= Mikrosimulationsmodell) und das *Gruppen- bzw. Regionssimulationsmodell*, für das in Abschnitt 3 bestimmte Eigenschaften abgeleitet wurden. Beide Modelle wurden auf die Stadt Werne (Kreis Unna, Nordrhein-Westfalen) angewandt, die am 31. 12. 1980 27 768 Einwohner hatte.

Für die Simulation der Bevölkerungsentwicklung haben wir die altersspezifischen Geburtenziffern des Jahres 1981 berechnet und im gesamten Prognosezeitraum (bis zu 90 Jahre) konstant gehalten. Für die Schätzung der Wernianer Sterbeziffern reichte die Fallzahl der Sterbefälle bei weitem nicht aus: Im Mittel kamen im Jahr 1981 nur 2 Sterbefälle auf jedes Altersjahr (Schaubild 3). Da auch bei größeren Regionen das Ausmaß der regionalen Sterblichkeitsunterschiede durchaus ungeklärt ist, weil diese Ziffern stark von der gewählten Art der Standardisierung abhängen, auf der der Regionsvergleich basiert (Birg, 1981), wurde hier auf die Berücksichtigung regionaler Besonderheiten bei den Sterbewahrscheinlichkeiten verzichtet. Grundlage für die Simulationsrechnung-

Schaubild 3
Sterbeziffern Bundesrepublik 1981 und Werne 1981 (weiblich)



gen, bei denen aus den o. a. Gründen zunächst nur die natürliche Bevölkerungsentwicklung (ohne Wanderungen) betrachtet wurde, sind also in beiden Modellen die gleichen altersspezifischen Geburtenziffern der Stadt Werne im Jahr 1981 bzw. die altersspezifischen Sterbeziffern des Bundesgebiets von 1981. Die für die Anwendung des Regions- bzw. Gruppenmodells ermittelten Varianzen der Geburten- und Sterbeziffern sind in *Tabelle 1* angegeben.

Der Unterschied zwischen beiden Simulationsmodellen besteht in der Art der Berücksichtigung des stochastischen Einflusses. Im *Einzelsimulationsmodell* werden die altersspezifischen Sterbeziffern auf jede Person der Wernianer Stadtbevölkerung einzeln angewandt. Zur Ermittlung der Geburtenzahl wurde aus dem Intervall (0,1) eine *gleichverteilte* Zufallszahl gezogen und mit dem Wert der Geburtenziffer für Frauen des Alters a verglichen. War die Zufallszahl kleiner als der Wert der Geburtenziffer f_a^w , wurde die Ziehung als Realisation eines Bernoulli-Experiments mit dem Ausgang „Geburt“ interpretiert, d. h. es wurde eine Geburt registriert (s. o.). Analog wurde beim Ziehen von Zufallszahlen für die Sterbeziffern verfahren: War die gezogene Zahl kleiner als der Wert der Sterbeziffer für ein bestimmtes Alter und Geschlecht, so wurde die Ziehung als Realisation eines Sterbefalles interpretiert und die betreffende Person aus dem Bevölkerungsbestand ausgeschieden.

Tabelle 1
 Varianzen der Prognoseparameter für das Regions- bzw. Gruppensimulationsmodell

Parameter		Standardabweichung	
Fertilität:	altersspezifische Geburtenziffern für die Stadt Werne im Jahr 1981, geglättet mit der Coale-Trussel-Funktion	geschätzt über alle Alter nach Gleichung (25): = 0,1925	
Mortalität:	Sterbeziffern der Bundesrepublik insgesamt 1981	getrennt geschätzt für zwei Altersgruppen nach Gleichung (25)	
		Alter 1- u. 29	Alter 29- u. 90
	Männer ...	2,14873	0,60309
	Frauen ...	1,00000	0,67319

Beim *Gruppen- bzw. Regionssimulationsmodell* wurden ebenfalls Zufallszahlen gezogen. Anders als beim Einzelsimulationsmodell wurde hier aber nicht von gleichverteilten Zufallszahlen ausgegangen, sondern von *normalverteilten*, wobei die Varianz der unterstellten Normalverteilung so gewählt wurde, daß sie dem Ausmaß der Differenz zwischen den „wahren“ Geburtenziffern der Stadt und den zufallsbedingten beobachteten Geburtenziffern in den verschiedenen Kalenderjahren entspricht (Varianzschätzung nach Gleichung 25). Die Abweichung der temporären zufallsabhängigen Geburtenziffern in Werne von den „wahren Werten“ von Werne wird durch den Zufallsterm in Gleichung (16.1) repräsentiert. Gleichung (16.1) ergibt mit Definition (18) und Annahme (21) folgende Gleichung zur Ermittlung der Geburtenziffern in den einzelnen Prognosejahren $t = 1982, 83, \dots$

$$(26) \quad f_{t,a}^w = f_{0,a}^w + f_{0,a}^w z_{t,a} \quad \text{für alle } a \text{ und } t \text{ im Prognosezeitraum.}$$

In Gleichung (26) wird für $z_{t,a}$ die gezogene normalverteilte Zufallszahl mit Mittelwert 0 und Varianz $\hat{\sigma}_{81}^2$ eingesetzt. Mit der aus Gleichung (26) resultierenden Geburtenziffer für das Alter a im Jahr t wird schließlich die Zahl der Frauen des gleichen Alters zur Ermittlung der Geburtenzahl multipliziert. Analog wurde bei der Berechnung der Sterbefälle verfahren. Anders als beim Einzelsimulationsmodell werden also beim Regions- oder Gruppensimulationsmodell die altershomogenen Gruppen als Gruppen fortgeschrieben, wodurch sich der Rechenaufwand im Vergleich zum Einzelsimulationsmodell erheblich reduziert. Für die Zufallsgröße z_t in (26) wurde eine Standardabweichung von 0,1925 geschätzt. Obwohl diese Standardabweichung als konstant und für alle Alter a gleich angenommen wird, ist die gezogene Zufallszahl natürlich für jedes Alter und für jedes Prognosejahr verschieden. Jedem einzelnen Simulationslauf entspricht somit eine spezifische Sequenz von gezogenen Zufallszahlen.

Bei den Sterbeziffern wurde die Standardabweichung der Ziffer der Stadt Werne von den Sterbeziffern der Bundesrepublik insgesamt als Maß für die Varianz der gezogenen Zufallszahl zugrunde gelegt. Die Varianzschätzung wurde analog zu Gleichung (25) durchgeführt, wobei allerdings für die Altersgruppen 1 bis unter 29 und für die Gruppen 29 bis unter 90 getrennte Schätzungen vorgenommen wurden, weil die Varianz bei den Sterbeziffern wegen der geringen Zahl der Sterbefälle stark vom Alter abhängt (s. Tabelle 1).

Sowohl beim Einzelsimulationsmodell als auch beim Regionssimulationsmodell hängen die Simulationsergebnisse natürlich von der Sequenz der gezogenen Zufallszahlen ab. Um zu allgemeinen Aussagen über die stochastische Variabilität zu gelangen, müssen möglichst viele alternative Simulationsläufe durchgeführt werden. In der vorliegenden Untersuchung wurden Gruppen von bis zu 120 Simulationsläufen pro Fallgruppe gerechnet. Aus den Prognose- bzw. Simulationsergebnissen der verschiedenen Läufe wurden Variationskoeffizienten ermittelt, definiert als Verhältnis der durchschnittlichen Abweichung der prognostizierten Bevölkerungszahl der Simulationsläufe vom Mittelwert der prognostizierten Bevölkerungszahlen. Der Variationskoeffizient wurde getrennt für die einzelnen Altersjahre und getrennt für Prognosezeiträume unterschiedlicher Länge berechnet, wobei zusätzlich nach dem Geschlecht differenziert wurde. Die feinste Differenzierung nach Alter, Geschlecht, Prognosedauer und Zahl der Simulationsläufe (m) ergibt:

$$VK_{t,a}^g = \frac{\sqrt{(1/m) \sum_j (B_{t,a}^g(j) - \bar{B}_{t,a}^g)^2}}{\bar{B}_{t,a}^g}$$

Hier ist t das Prognosejahr, j die Nummer des Simulationslaufs, g das Geschlecht und m die Zahl der Läufe. Um den Einfluß der Zahl der Läufe auf den Variationskoeffizienten zu simulieren, wurden alternative Simulationen in Abhängigkeit von $m = 1, 2, \dots, 120$ durchgeführt.

Bei kleinen Bevölkerungen treten bei der Berechnung der Zahl der Geburten und Sterbefälle Zahlen auf, die kleiner als 1 sein können. In Werne lebten am 31. 12. 1980 beispielsweise 235 Männer im Alter von 20 bis unter 21 Jahren. Multipliziert man diese Zahl mit der Sterbewahrscheinlichkeit der 20jährigen Männer ($= 0,00162$), so erhält man 0,381 Sterbefälle, eine Zahl, die nach dem Runden Null ergibt. Um eine systematische Unterschätzung der Sterbefälle, die durch das Runden auf ganzzahlige Werte entstehen würde, zu vermeiden, wurden sämtliche Simulationsrechnungen mit reellwertigen Variablen durchgeführt.

Werden für jedes der beiden Modelle eine alternative Anzahl von m Simulationsläufen durchgeführt, so hängt der Variationskoeffizient von folgenden Faktoren ab, die in den beiden Modellen ein unterschiedliches Gewicht haben:

- (1) Alter, Geschlecht,
- (2) Länge des Prognosezeitraums,

- (3) Gliederungsstruktur der Bevölkerung (= Differenziertheit der Gliederung nach Alter und Geschlecht),
- (4) Größe der Bevölkerung,
- (5) Zahl der Simulationsläufe.

Die Aussagen über die stochastische Variabilität der Prognosen müssen im folgenden nach diesen Dimensionen differenziert werden:

(1) Die Abhängigkeit des Variationskoeffizienten vom Alter

Für diejenigen Altersjahrgänge, die am Beginn des Prognosezeitraums bereits leben, sind die Variationskoeffizienten wesentlich kleiner als für diejenigen Gruppen, die im Prognosezeitraum geboren werden. Die prognostizierten Bestandszahlen der am Ende eines Prognosezeitraums von 30 Jahren lebenden Personen, die dann jünger als 30 Jahre sind, unterliegen drei verschiedenen stochastischen Einflüssen: (a) Die Zahl der Frauen im gebärfähigen Alter hängt von der Stochastik des Sterbeprozesses ab, (b) die Zahl der von diesen Frauen geborenen Kinder hängt von der Stochastik des Geburtenprozesses ab, (c) die Zahl der geborenen Kinder, die bis zum Prognosezeitpunkt überleben,

Schaubild 4:
Variationskoeffizienten nach 30 Prognosejahren (Gruppensimulation)

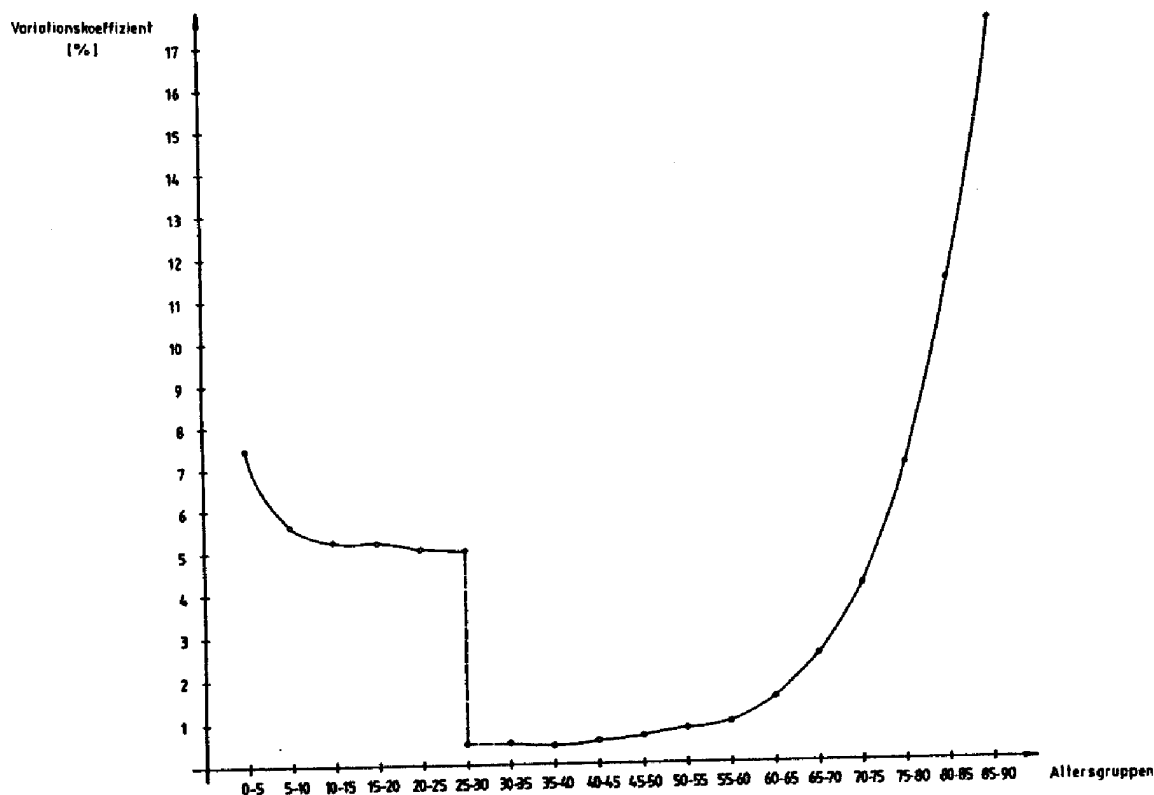
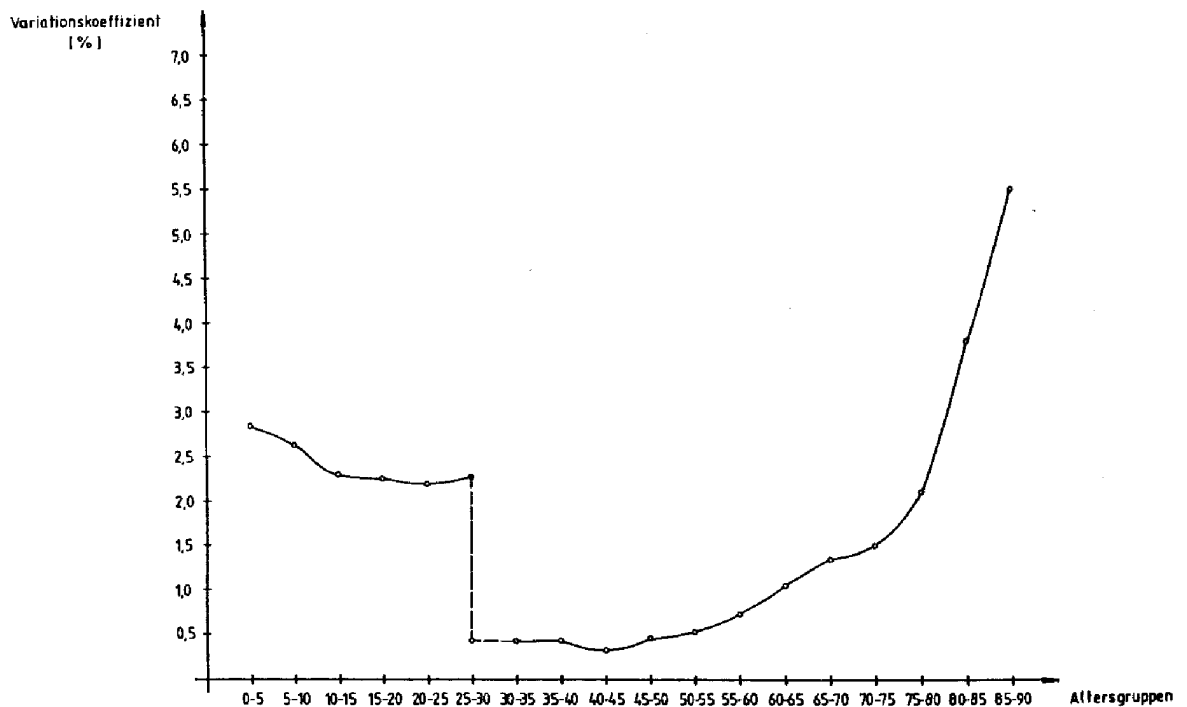


Schaubild 5:
Variationskoeffizienten nach 30 Prognosejahren (Einzelsimulation)



hängt wiederum von der Stochastik des Sterbeprozesses ab. Je jünger die Personen in der Gruppe der unter 30jährigen am Ende des 30jährigen Prognosezeitraums sind, desto später werden sie geboren und desto mehr macht sich die Stochastik bei der Prognose der Zahl der Frauen bemerkbar, die in der Altersklasse der potentiellen Mütter sind. Deshalb haben die Kurven in Figur 4 (Regions- bzw. Gruppenmodell) und Figur 5 (Einzelmodell) zunächst einen fallenden Verlauf. Danach steigen die Kurven wieder an, weil die Sterbeziffern und damit auch die Varianzen der Sterbefälle mit zunehmendem Alter rasch ansteigen und zusätzlich die Besetzungszahlen sinken, so daß der Variationskoeffizient als Quotient aus Standardabweichung und durchschnittlicher Besetzungszahl zunimmt. Es sei hier angemerkt, daß die Kurven in den Figuren 4 und 5 (sowie die Kurven in den anderen Schaubildern) nicht zur Ableitung der Varianz für einzelne Altersjahre herangezogen werden dürfen; die Verbindungslinien zwischen den Punkten für die 5-Jahres-Gruppen dienen nur der Anschaulichkeit.

(2) Die Abhängigkeit des Variationskoeffizienten von der Länge des Prognosezeitraums

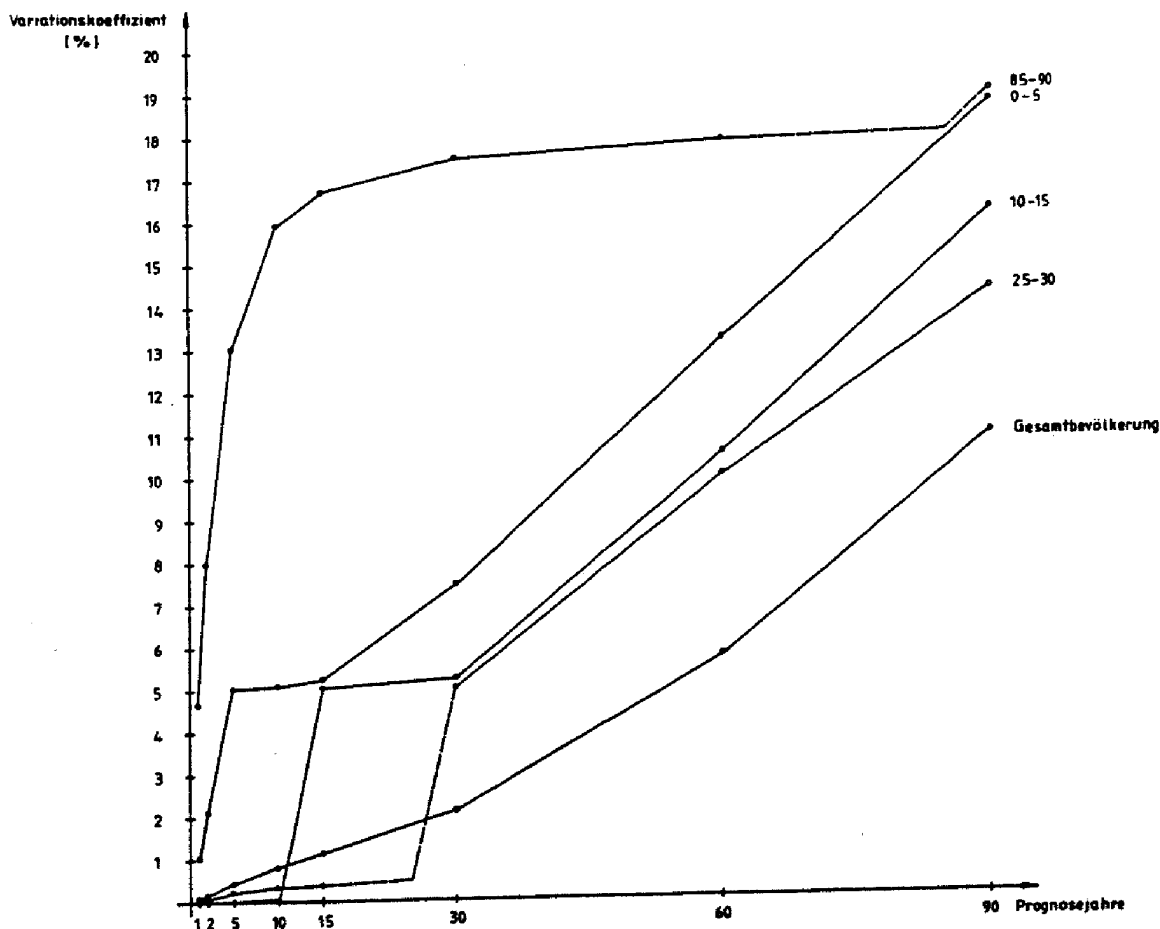
Der Variationskoeffizient steigt in beiden Modellen mit der Länge des Prognosezeitraums tendenziell an, doch gibt es systematische Unterschiede zwischen den Modellen. Im Gruppensimulationsmodell ist der Variationskoeffizient um

ein Vielfaches größer als im Einzelsimulationsmodell. Tendenziell gilt für beide Modelle: Je länger der Prognosezeitraum, desto weiter öffnet sich die Schere:

	Variationskoeffizient in % in Abhängigkeit von der Prognosedauer		
	15 Jahre	30 Jahre	60 Jahre
Gruppensimulationsmodell	1,13	2,11	5,71
Einzelsimulationsmodell	0,30	0,41	0,84

Die Gruppe der 10- und 15jährigen umfaßt bis zu einer Prognosedauer von 10 Jahren nur Personen, die bereits geboren sind. Mit dem 10. Prognosejahr kommt ein Jahrgang in diese Gruppe hinein, der während des Prognosezeitraums geboren wird. Die Kurve hat daher im 10. Prognosejahr eine Sprungstelle, von der ab der Anstieg als Variationskoeffizient steiler verläuft, so lange neue Jahrgänge zur Gruppe hinzukommen, d. h. bis zum 15. Prognosejahr. Danach verläuft der Anstieg wieder flacher. Dieser gebrochene Verlauf ist den Kurven in beiden Modellen gemeinsam (Schaubilder 6 u. 7). Für die Stadt-

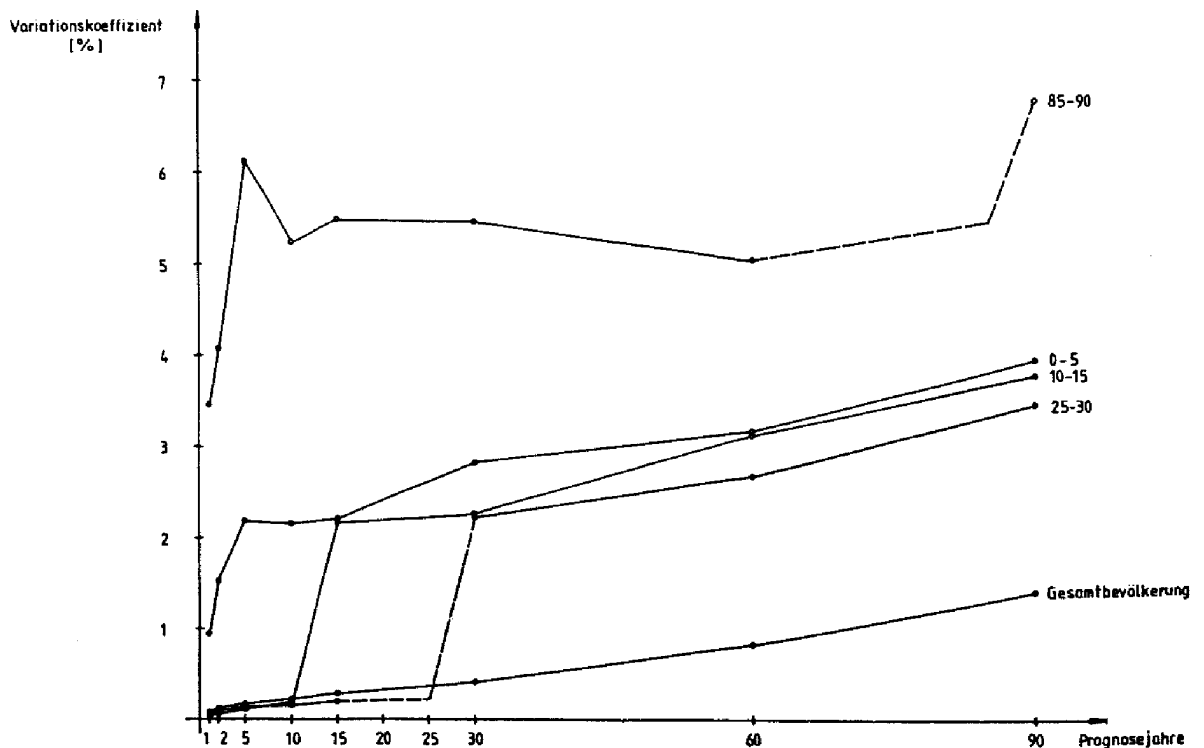
Schaubild 6:
Abhängigkeit des Variationskoeffizienten vom Prognosezeitraum (Gruppensimulation)



und Regionalplanung ist von Bedeutung, daß nicht mit einheitlichen Fehlerrisiken für alle Altersgruppen gerechnet werden darf. So ist beispielsweise die Bandbreite für die Zahl der über 65jährigen (Krankenhausplanung) rd. dreimal so groß wie für die unter 15jährigen (Schulplanung).

Zwei Besonderheiten seien hervorgehoben. Für das Gruppensimulationsmodell steigt in Figur 6 die Kurve für die 85—90jährigen zunächst steil an und wird wieder flacher, weil die Personen in dieser Gruppe der Stochastik des Sterbeprozesses um so länger ausgesetzt sind, je größer der Prognosezeitraum ist. Aber je länger die Personen in der Gruppe 85—90 dem Sterbeprozess ausgesetzt sind, mit desto jüngeren Jahren treten sie in diese Gruppe ein und desto kleiner sind die Sterbeziffern beim Eintritt in die Gruppe. Der Effekt kleiner Sterbeziffern überwiegt schließlich, so daß die Kurve nach 30 Jahren nur noch schwach ansteigt. Im Einzelsimulationsmodell wird diese Regelmäßigkeit von einem Faktor, der mit der Logik des Einzelsimulationsmodells zusammenhängt, im vorliegenden Fallbeispiel so stark überlagert, daß er nicht mehr erkennbar ist: Im Einzelsimulationsmodell *sinkt* der Variationskoeffizient für die Gruppe der 85—90jährigen mit zunehmender Prognosedauer (Figur 7). Dies liegt an einer modellstrukturbedingten Eigenschaft des Einzelsimulationsmodells, die darin besteht, daß der Variationskoeffizient der Binomialverteilung mit wachsender Besetzungszahl n abnimmt:

Schaubild 7:
Abhängigkeit des Variationskoeffizienten vom Prognosezeitraum (Einzelsimulation)



$$VK = \sqrt{\frac{1-p}{np}}$$

Da die Besetzungszahl n in der Gruppe der 85–90jährigen vom 10. bis zum 60. Prognosejahr ansteigt, hat die Kurve in diesem Zeitraum einen fallenden Verlauf.

Die starke Abhängigkeit der Fehlervariabilität von der Bevölkerungszahl ist eine allgemeine Eigenschaft des Einzelsimulationsmodells, die man im Auge behalten muß, um Fehlinterpretationen zu vermeiden. Bei wachsenden Bevölkerungen kann der negative Effekt der Bevölkerungszahl auf den Variationskoeffizienten den positiven Effekt der Länge des Prognosezeitraums überkompensieren, mit der Folge, daß mit steigendem Prognosehorizont eine *sinkende* Fehlervariabilität verbunden ist — eine Eigenschaft des Einzelsimulationsmodells, die zu einer negativen Beurteilung dieses Modells im Vergleich zum Gruppensimulationsmodell führen muß, bei dem dieser Effekt nicht auftritt.

(3) Der Einfluß der Gliederungsstruktur auf den Variationskoeffizienten

Sämtliche Simulationsläufe wurden für eine feine Altersgliederung nach einzelnen Altersjahren durchgeführt, daher lassen sich die Prognoseergebnisse zu alternativen Altersgliederungen zusammenfassen. Tendenziell gilt für beide Modelle, daß der Variationskoeffizient um so kleiner ist, je größer die Altersgruppen sind, zu denen die Einzelergebnisse zusammengefaßt werden. Während aber dieser Effekt im Gruppensimulationsmodell relativ klein ist, hat er im Einzelsimulationsmodell eine so überragende Bedeutung, daß Aussagen über die stochastische Variabilität für diesen Modelltyp, zu dem die meisten Mikrosimulationsmodelle gehören, *inhaltsleer* sind, wenn der Variationskoeffizient nur für das Gesamttaggregat angegeben wird. In *Tabelle 2* ist zu erkennen, wie stark sich die Variabilität reduziert, wenn von einer Gliederung in einzelne Altersjahre zu einer Gliederung in 5er-Altersgruppen übergegangen wird: Beim Einzelsimulationsmodell halbieren sich die Variationskoeffizienten, beim Gruppenmodell sinken sie nicht wesentlich. Im Einzelsimulationsmodell haben die Variationskoeffizienten für die nach einzelnen Altersjahren abgegrenzten Bevölkerungsbestände nach 15 Prognosejahren Werte im Intervall zwischen 0,46 % (25jährige) und 16,00 % (90jährige); ihr arithmetisches Mittel beträgt 3,2 % (*Tabelle 2*, Sp. 1). Der Variationskoeffizient für das Gesamttaggregat ist dagegen nur 0,30 % (Sp. 2) und liegt damit um den *Faktor 10* unter dem arithmetischen Mittel der Einzelwerte. Beim Gruppensimulationsmodell beträgt das arithmetische Mittel der Einzelkoeffizienten 3,78 %, der Variationskoeffizient für das Gesamttaggregat ist 1,13 %, d. h. der Effekt der Gliederungsstruktur ist hier wesentlich kleiner.

Tabelle 2

Der Einfluß der Altersgliederung auf den Variationskoeffizienten der Prognoseergebnisse im Einzelsimulationsmodell und im Regions- bzw. Gruppensimulationsmodell (Prognosezeitraum = 15 Jahre)

Alter	Einzelsimulationsmodell		Regions- bzw. Gruppensimulationsmodell	
	Arithmetisches Mittel der einzelnen VK in der Altersgruppe	VK der Gruppe als Ganzes	Arithmetisches Mittel der einzelnen VK in der Altersgruppe	VK der Gruppe als Ganzes
	— in % —			
0— 5	5,140	2,221	5,251	5,226
5—10	4,972	2,164	5,112	5,089
10—15	4,999	2,193	5,044	5,022
15—20	0,533	0,253	0,171	0,107
20—25	0,530	0,233	0,334	0,302
25—30	0,503	0,200	0,407	0,382
30—35	0,502	0,234	0,442	0,413
35—40	0,583	0,265	0,403	0,358
40—45	0,697	0,378	0,443	0,385
45—50	0,994	0,469	0,592	0,499
50—55	1,246	0,558	0,818	0,712
55—60	1,401	0,632	1,065	0,884
60—65	2,007	0,877	1,627	1,433
65—70	2,535	1,182	2,575	2,323
70—75	3,510	1,633	4,189	3,811
75—80	6,428	2,641	7,176	6,226
80—85	8,018	3,883	12,122	10,715
85—90	13,007	5,474	20,284	16,690
0— 3	5,259	3,008	5,267	5,246
3— 6	4,919	3,007	5,225	5,206
6—10	5,009	2,515	5,085	5,064
10—16	4,270	1,851	4,236	4,274
16—19	0,511	0,315	0,138	0,101
19—25	0,526	0,227	0,319	0,283
15—65	0,899	0,146	0,613	0,289
65 u.ä.	6,699	0,939	9,269	3,876
0—90	3,200	0,299	3,781	1,130

(4) *Der Einfluß der Bevölkerungszahl (Regionsgröße) auf den Variationskoeffizienten*

Der Einfluß der Bevölkerungszahl auf die stochastische Variabilität folgt im Einzelsimulationsmodell tendenziell dem Quadratwurzelgesetz: Eine Verdopplung der Bevölkerungszahl führt zu einer Halbierung des Variationskoeffizienten. Die-

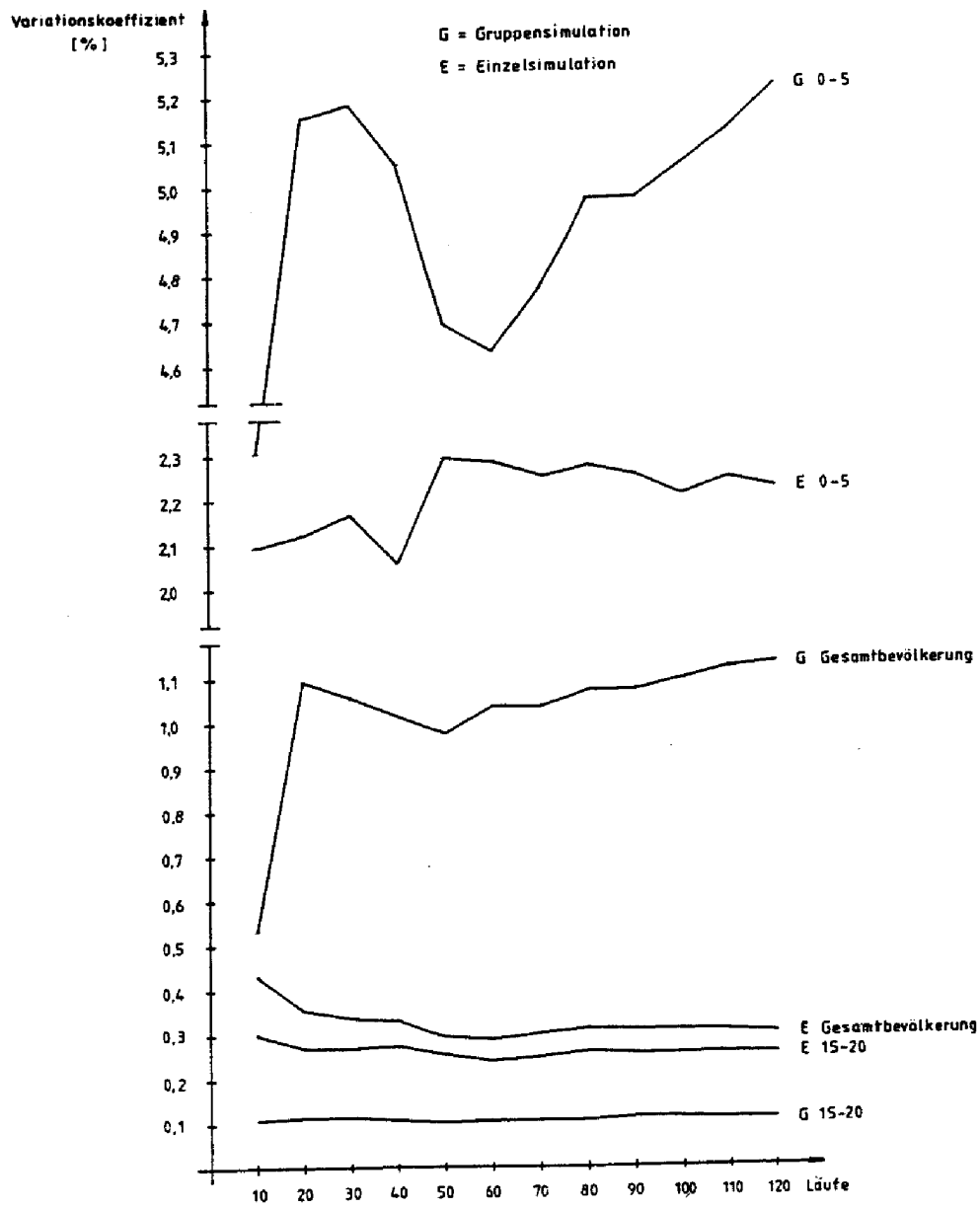
se von Galler (1983, S. 167 f.) für Mikrosimulationsmodelle abgeleitete tendenzielle Regel läßt sich für die *Geburten* und *Sterbefälle* in einem bestimmten Jahr t unmittelbar aus der Definition des Variationskoeffizienten ableiten, nicht jedoch für den Bevölkerungsbestand nach einer großen Zahl von Prognosejahren. Wir haben die Regel geprüft, indem wir zusätzlich zu den Simulationsrechnungen für die Stadt Werne auch für einen Stadtteil von Werne (Stockum), der rd. 4700 Einwohner hat, Simulationsrechnungen auf der Basis des Einzelsimulationsmodells durchführten. Auf der Ebene der 5er-Altersgliederung betrug die größte Abweichung zwischen den originären Simulationsergebnissen für den Stadtteil Stockum und den Ergebnissen nach dem Quadratwurzelgesetz 17 %; in 9 der 18 Altersgruppen lag sie unter 5 %.

Für das Gruppensimulationsmodell konnten Simulationsrechnungen für den Stadtteil Stockum in Ermangelung von Daten zur Schätzung der Geburtenziffern in diesem Stadtteil nicht durchgeführt werden. Weitere Simulationsrechnungen auf der Basis des Regions- bzw. Gruppensimulationsmodells sind für die Analyse des Einflusses der Bevölkerungszahl (Regionsgröße) auf den Fehlerspielraum regionaler Bevölkerungsprognosen um so nötiger, als eine Beurteilung dieses Einflusses, die sich allein auf das Einzelsimulationsmodell stützt, nicht aussagekräftig ist. Die tendenzielle Gültigkeit des Quadratwurzelgesetzes für Einzelsimulationsmodelle ist keine brauchbare Grundlage für die Schätzung des Einflusses der Regionsgröße, weil in diesem Modell die Geburtenziffern nicht in eine regionstypische und in eine zufallsabhängige (temporär unterschiedliche) Komponente zerlegt werden. Im Einzelsimulationsmodell wird nur eine Komponente berücksichtigt, und die stochastische Variabilität der Ergebnisse resultiert aus der Variabilität des Geburten- und des Sterbeprozesses, die mit wachsendem n stark sinkt, so daß der Prognosefehler für bestimmte Altersgruppen mit steigender Prognosedauer, wie schon gezeigt, *abnehmen* kann, wenn die Besetzungszahl in der Altersgruppe wächst. Bei großen Regionen führt das Einzelsimulationsmodell daher tendenziell zu einer *Unterschätzung* des Fehlerspielraums. Für die Stadt Werne waren die Fehlerspielräume in beiden Modellen etwa gleich, aber nur dann, wenn man die nach einzelnen Altersjahren differenzierten Variationskoeffizienten vergleicht. Bei größeren Altersgruppen oder gar bei der Bevölkerung insgesamt macht sich der Effekt der Besetzungszahlen im Einzelsimulationsmodell so stark bemerkbar, daß die Variationskoeffizienten des Einzelsimulationsmodells deutlich kleiner sind als die des Regions- bzw. Gruppensimulationsmodells (Tabelle 2, Spalten 2 und 4).

(5) *Abhängigkeit der Variationskoeffizienten von der Zahl der Simulationläufe*

Die Mittelwerte der Simulationsergebnisse des Gruppensimulationsmodells und des Einzelsimulationsmodells unterscheiden sich wesentlich weniger voneinander als ihre Varianzen bzw. ihre Variationskoeffizienten. Die Unterschiede der Mittelwerte verringern sich mit der Zahl der Simulationläufe (Figur 8), während

Schaubild 9:
Variationskoeffizienten nach 15 Prognosejahren



die Unterschiede hinsichtlich des Variationskoeffizienten (im vorliegenden Fallbeispiel) nicht geringer werden, wenn die Zahl der Simulationsläufe erhöht wird (Figur 9). Für die Beurteilung der Unterschiede zwischen den Mittelwerten hätten schon 20 Simulationsläufe ausgereicht, aber nicht für die Analyse der Variationskoeffizienten.

Für die meisten Altersgruppen liegt der Variationskoeffizient des Gruppensimulationsmodells über dem des Einzelsimulationsmodells, für die Altersgruppen 0–5 und für die Gesamtbevölkerung um mehr als das Doppelte (Figur 9), aber für die Gliederung nach einzelnen Altersgruppen sind die Koeffizienten annähernd gleich (Tabelle 2). Dies zeigt, daß generalisierende Aussagen von der Art, daß die stochastische Variabilität im Gruppensimulationsmodell größer ist als im Einzelsimulationsmodell, nur unter Einschränkungen möglich sind. Immerhin läßt sich für das vorliegende Fallbeispiel festhalten, daß für das Gesamttaggregat der Bevölkerung auf der Basis eines Prognosezeitraums von 15 Jahren nach dem Einzelsimulationsmodell eine kleinere stochastische Prognosevariabilität errechnet wird als nach dem Gruppenmodell. Die Variabilität des Einzelsimulationsmodells sollte daher allenfalls als *unterer* Schätzwert für die (unbekannte) wahre Variabilität verwendet werden.

5. Resümierende Schlußfolgerungen

Die Variabilität von Bevölkerungsprognosen beruht auf drei verschiedenen Fehlerquellen, auf Fehlern der Ausgangsdaten, auf den Abweichungen der Prognoseannahmen über Ursache-Wirkungsbeziehungen von den realen Zusammenhängen (Annahmenfehlern) sowie auf der grundsätzlich stochastischen Eigenschaft des Bevölkerungsprozesses. In der Praxis haben die Annahmenfehler das größte Gewicht, bei regionalen Bevölkerungsprognosen vor allem die Annahmenfehler bezüglich der Wanderungen.

Die stochastische Variabilität der Prognosen, die aus dem stochastischen Charakter der realen Entwicklung resultiert, hängt von der Größe der Region ab (Einwohnerzahl). Für das Fallbeispiel der Stadt Werne (Größenklasse um 30 000 Einwohner) wurde die stochastische Variabilität auf der Basis eines Modells ohne Wanderungen für zwei verschiedene Modelltypen untersucht, und zwar für ein Einzelsimulationsmodell und für ein Regions- bzw. Gruppensimulationsmodell. Während die Erwartungswerte der prognostizierten bzw. simulierten Bevölkerungszahl — gegliedert nach Alter und Geschlecht — in beiden Modellen weitgehend übereinstimmen, ist ihre stochastische Variabilität, gemessen durch den Variationskoeffizienten der Prognoseergebnisse aus einer großen Zahl von Simulations- bzw. Prognoseläufen, unterschiedlich: Die stochastische Variabilität des Einzelsimulationsmodells ist in dem vorliegenden Fall wesentlich kleiner als die des Gruppen- bzw. Regionssimulationsmodells. Der auf der Basis des Einzelsi-

mulationsmodells geschätzte Fehlerspielraum sollte daher allenfalls als *unterer Grenzwert* für den wahren Fehlerspielraum interpretiert werden.

Das Ausmaß der stochastischen Variabilität hängt im Einzelsimulationsmodell stark von der Feinheit der Altersgliederung ab. Allgemeine Aussagen über die Ex-ante-Fehlervariabilität dieses Modells sind somit problematisch. Denn es läßt sich meist eine Altersgliederung finden, für die der Fehlerspielraum klein ist. Im Gegensatz dazu sind die Effekte der Altersstruktur im Regions- bzw. Gruppensimulationsmodell kleiner.

Aus inhaltlicher Sicht hat das Regions- bzw. Gruppenmodell den Vorteil, daß es regions- bzw. gruppenspezifische Verhaltensweisen berücksichtigt. In diesem Modell wird die altersspezifische Geburtenwahrscheinlichkeit jedes Individuums in einen kollektiven d. h. regions- bzw. gruppentypischen Teil und in eine individuell bedingte personenabhängige Komponente zerlegt. Dadurch lassen sich die intertemporalen Schwankungen der regions- bzw. gruppenspezifischen Geburtenziffern als stochastische Realisationen der Ereignisse auf Individualebene interpretieren und im Rahmen von Simulations- bzw. Prognoseberechnungen zur Schätzung der stochastischen Variabilität nutzen. Die Ergebnisse ermöglichen Aussagen über den zu erwartenden Fehlerspielraum für Gemeinden der Einwohnerklasse um 30 000, und zwar für jene Variabilität des Prognosefehlers, die aus der natürlichen Bevölkerungsentwicklung (ohne Wanderungen) resultiert. Die Berechnungen sollen auch für Städte bzw. Regionen anderer Einwohnergrößenklassen durchgeführt werden. Als wichtiges Zwischenergebnis kann festgehalten werden, daß die Stadt- und Regionalplanung nicht mit einheitlichen Fehlerspielräumen für alle planungsrelevanten Altersgruppen rechnen darf. So liegt z. B. der Fehlerspielraum (Variationskoeffizient) für die Zahl der Einwohner in der Altersgruppe über 65 Jahre (Krankenhausplanung, Altenhilfe usw.) bei einem Prognosezeitraum bis zu 30 Jahren in dem vorliegenden Beispiel beim Dreifachen des Fehlerspielraums für die unter 15jährigen (Schulplanung). Wichtig ist, daß sich sämtliche Fehlerspielräume beträchtlich erhöhen, wenn der Wanderungsprozeß zusätzlich zum Geburten- und Sterbeprozess in die Analyse einbezogen wird.

Literatur

- Birg, H., Maneval, K., Masuhr, K., 1979: Synopse von Verfahren zur regionalen Bevölkerungs- und Arbeitsplatzprognose im Bereich des Bundes und der Bundesländer. Gutachten im Auftrag des Bundesministers für Raumordnung, Bauwesen und Städtebau, des Bayerischen Staatsministers für Landesentwicklung und Umweltfragen, des Niedersächsischen Ministers des Inneren, der Staatskanzlei Rheinland-Pfalz und der Staatskanzlei des Landes Nordrhein-Westfalen, Berlin/München/Basel (nicht publiziert).
- Birg, H., 1979: Zur Interdependenz der Bevölkerungs- und Arbeitsplatzentwicklung — Grundlagen eines simultanen interregionalen Modells für die Bundesrepublik Deutschland. Sonderheft Nr. 131 des Deutschen Instituts für Wirtschaftsforschung, Duncker und Humblot, Berlin.

- Birg, H., 1980: Berechnungen zur langfristigen Bevölkerungsentwicklung in den 343 kreisfreien Städten und Landkreisen der Bundesrepublik Deutschland, in: Vierteljahrshefte zur Wirtschaftsforschung, Heft 2, Berlin.
- Birg, H., 1982: Regionale Mortalitätsunterschiede in der Bundesrepublik Deutschland — Ein Problemaufriß, in: Materialien des Instituts für Bevölkerungsforschung und Sozialpolitik, Bd. 4, Bielefeld.
- Birg, H., 1984: Argumente für die Verwendung von Mehr-Regionen Modellen bei der Erstellung von Bevölkerungsprognosen, in: Informationen zur Raumentwicklung, H. 12, Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn.
- Birg, H., 1986: Regionale Demographie und Regionalwissenschaftliche Analyse, in: Materialien des Instituts für Bevölkerungsforschung und Sozialpolitik, Bd. 21, Bielefeld.
- Bucher, H., 1984: Das regionale Bevölkerungsprognosemodell der BfLR, in: Informationen zur Raumentwicklung, H. 12, Bundesforschungsanstalt für Landeskunde und Raumordnung, Bonn.
- Chiang, Long Chin, 1984: The Life Table and its Applications, Malabar, Florida.
- Coale, A. J., Trussel, T. J., 1974: Model Fertility Schedules: Variations in the Age Structure of Childbearing in Human Populations, in: Population Index, Nr. 40, Princeton, S. 185—203.
- Coale, A. J., Trussel, T. J., 1978: Technical Note: Finding the two Parameters that Specify a Model Schedule of Marital Fertility, in: Population Index, Nr. 44, Princeton, S. 203—213.
- Feichtinger, G., 1971: Stochastische Modelle demographischer Prozesse, in: Lecture Notes in Operation Research and Mathematical Systems, Vol. 44, Springer Verlag, Berlin/Heidelberg/New York.
- Galler, H. P., 1983: Mikrosimulationsmodelle als demographische Planungsgrundlage, in: Birg, H. (Hrsg.), Demographische Entwicklung und gesellschaftliche Planung, Forschungsberichte des Instituts für Bevölkerungsforschung und Sozialpolitik, Bd. 6, Frankfurt/New York.
- Gatzweiler, H.-P., Stiens, G., 1982: Regionale Mortalitätsunterschiede in der Bundesrepublik Deutschland — Daten und Hypothesen, in: Jahrbuch für Regionalwissenschaft 3.
- Hoem, J. M., 1973: Levels of Error in Population Forecasts, Artikler fra Statistisk Sentralbyra, Nr. 61, Oslo.
- Keyfitz, N., 1985: A Probability Presentation of Future Population, in: Zeitschrift für Bevölkerungswissenschaft, H. 2, Boppard/Rh.
- Koch, H., 1986: Bayern regional 2000, Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen (Hrsg.), München.
- Orcutt/Caldwell/Wertheimer, 1976: Policy Exploration Through Microanalytic Simulation, The Urban Institute, Washington, D. C.
- Pollard, J. H., 1966: On the Use of the Direct Matrix Product in Analysing Certain Stochastic Population Models, in: Biometrika, Nr. 53, S. 397—415.
- Pollard, J. H., 1973: Mathematical Models for the Growth of Human Populations, Cambridge University Press, Cambridge.
- Steger, A., 1980: Haushalte und Familien bis zum Jahr 2000. Eine mikroanalytische Untersuchung für die Bundesrepublik Deutschland, in: Schriftenreihe des Sonderforschungsbereichs 3 der Universitäten Frankfurt und Mannheim, Mikroanalytische Grundlagen der Gesellschaftspolitik, Bd. 3, Campus-Verlag, Frankfurt/New York.
- Sykes, Z. M., 1969: Some Stochastic Version of the Matrix-Model for Population Dynamics, in: Journal of the American Statistical Association, Nr. 64, S. 111—130.