Kapitel 6

Modellierung und Simulation von Bausparkollektiven

Seit einigen Jahren besteht eine enge Zusammenarbeit zwischen den Landesbausparkassen (LBS) und unserer Arbeitsgruppe am ZAIK. Die Schwerpunkte der Kooperation bestehen sowohl in der Analyse großer Datenmengen, als auch in der Simulation komplexer sozioökonomischer Systeme. Die Ergebnisse werden von den Landesbausparkassen für ihre langfristige Geschäftsplanung verwendet.

Im Folgenden wird unser im Einsatz befindliches Simulationsmodell vorgestellt. Nach einem generellen Überblick werden die jüngsten Weiterentwicklungen des Modells beschrieben. In den beiden letzten Abschnitten werden zwei alternative Modellierungsansätze erläutert, die im Rahmen mehrerer Dissertationen entwickelt worden sind.

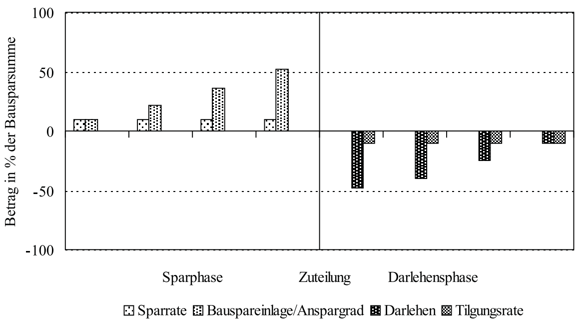

Abbildung 6.1: *Idealisierter* **Kontoverlauf** *mit konstanter Sparzahlung und konstanter Tilgung. Der Anspargrad ist der Anteil der Bauspareinlage an der Bausparsumme. Die Bewertungszahl (nicht abgebildet) ist ein abstraktes Maß für die bisher erbrachte Sparleistung.*

6.1 Ein deterministisches Modell für Bausparkollektive

6.1.1 Das Bausparen und unsere Tätigkeit – eine Übersicht

Der Ablauf eines Bausparvertrages gliedert sich für den individuellen Bausparer[1] im wesentlichen in zwei Phasen: Der Ablauf beginnt mit dem Abschluss eines Bausparvertrages über eine bestimmte Bausparsumme, die der Bausparer in sein Immobilienprojekt investieren will. In der Sparphase zahlt der Bausparer auf sein Konto ein und stellt seine Bauspareinlagen der Bausparkasse zur Verfügung. Mit der Zuteilung endet die Sparphase. Danach kann sich der Bausparer die Bausparsumme auszahlen lassen. Der Zeitpunkt der Zuteilung hängt davon ab, für welchen Zeitraum der Bausparer welchen Betrag der Kasse zur Verfügung gestellt hat. Die Differenz zwischen der Bausparsumme und der Bauspareinlage wird dem Bausparer als Darlehen von der Kasse zur Verfügung gestellt. Mit der Auszahlung der Bausparsumme beginnt die Darlehensphase. In der Darlehensphase tilgt der Bausparer sein Darlehen mit einer vorgegebenen Mindesttilgungsrate (Abb. **??**). Der Bausparer kann durch Aktionen wie vorzeitige Kündigung innerhalb der Sparphase, Erhöhung und Ermässigung der Bausparsumme, Fortsetzung der Sparphase und Sondertilgung des Darlehens den Prozess von seiner Seite aus steuern.

Die Gesamtheit aller Bausparer wird als Bausparkollektiv bezeichnet. Die Funktion der Bausparkasse in dem Prozess besteht darin, dass sie innerhalb des Bausparkollektivs die Bauspareinlagen der einen Bausparer zusammen-

[1] Zur Vereinfachung der Lesbarkeit unseres Artikels verwenden wir die männliche Form, wenn sowohl Frauen als auch Männer als Bausparer gemeint sind.

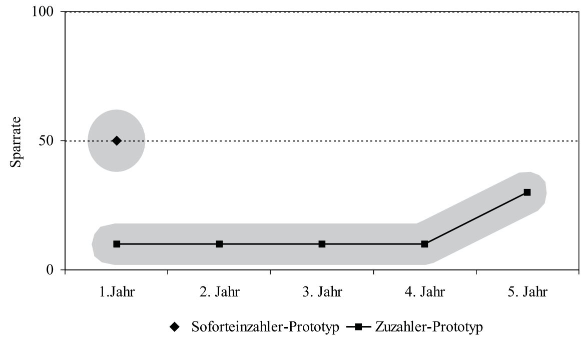

Abbildung 6.2: **Clusterung** *mit zwei Clustern (graue Flächen) und dazugehörenden Prototypen. Der Prototyp Soforteinzahler repräsentiert das Verhalten, mit einer einmaligen Zahlung sofort den Mindestanspargrad zu erreichen. Der Prototyp Zuzahler repräsentiert das Verhalten, mit einer Erhöhung der Sparzahlung die Sparphase zu beenden.*

fasst und als Darlehen an die anderen Bausparer verteilt. Für die Geschäftsplanung der Bausparkasse sind die Entwicklung der gesamten Bauspareinlagen, also die Überlagerung aller individuellen Einzahlungen, und die Entwicklung der zur Verfügung gestellten Darlehen, also die Überlagerung aller ausstehenden Darlehen, von zentraler Bedeutung. Unsere Arbeitsgruppe analysiert das Verhalten von Bausparkollektiven in der Vergangenheit und erstellt aufgrund der gewonnenen Ergebnisse Prognosen für die zukünftige Entwicklung der Kollektive. Außerdem erstellen wir auf der Basis vorgegebener Eckdaten der Bausparkassen Szenarien für die zukünftige Entwicklung von Bausparkollektiven. Dabei sind alle Problemebenen in die Arbeitsgruppe integriert: Die zugrunde liegenden statistischen Modelle und die darauf aufbauenden Algorithmen werden hier entwickelt. Anschließend werden die Algorithmen implementiert und auf die Kollektivdaten, welche die Bausparkassen uns bereitstellen, angepasst und angewendet. Auftretende Probleme mit den Daten und neue Anforderungen seitens der Bausparkasse fließen somit unmittelbar in die Weiterentwicklungen ein, welche auf den unterschiedlichen Probleminstanzen getätigt werden.

6.1.2 Modellansatz

Die statistische Darstellung des Verhaltens eines Bausparkollektives lässt sich in zwei Schritte unterteilen. Im ersten Schritt wird ausschließlich das Sparverhalten betrachtet. Der einzelne Bausparer kann seine Sparzahlungen frei gestalten, was zur Folge hat, dass die Menge aller Sparverhaltensweisen des Bausparkollektivs sehr inhomogen ist. Der statistische Zugang erfolgt über die Clusteranalyse: Zuerst werden diejenigen Verträge aus dem Bestand von Bausparverträgen herausgegriffen, von denen eine vollständige Sparphase vorliegt. In der Clusteranalyse werden die Sparraten, also die Sparzahlungen relativ zur Bausparsumme des jeweiligen Vertrages, der einzelnen Verträge miteinander verglichen. Die Verträge mit ähnlichem Sparverhalten werden zu einem Cluster zusammengefasst, dessen mittlere Sparrate, Prototyp genannt, ermittelt wird (Abb. **??**). Die absoluten Sparzahlungen des Clusters ergeben sich aus dem Produkt der Sparraten des Prototypen mit der Bausparsumme des Clusters.

Anschließend werden die übrigen Verträge, bei denen die Sparzahlungen noch nicht beendet sind, denjenigen Prototypen zugeordnet, die im Sparverhalten den Verträgen am ähnlichsten sind. Außerdem wird jedem Prototypen die Bausparsumme all derjenigen Verträge, die dem jeweiligen Prototypen zugeordnet wurden, zugewiesen. Der Prototyp repräsentiert sowohl das historische Sparverhalten der Gesamtheit der zugeordneten Verträge, als auch das zukünftige Sparverhalten der Gesamtheit (Abb. **??**).

Im zweiten Schritt wird das weitere Verhalten, das innerhalb des Vertragsablaufs auf das Sparverhalten folgt, in die statistische Darstellung des Bausparkollektivs miteinbezogen. Sowohl das Tilgungsverhalten als auch die anderen Aktionen, mit denen der Bausparer den Bausparprozess steuern kann, werden durch bestimmte Bausparaktionen beschrieben. Die Bausparaktionen spannen in ihrer

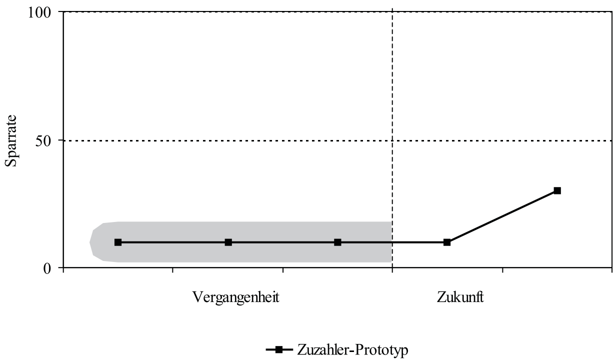

Abbildung 6.3: **Zuordnung** *von Verträgen mit noch nicht beendeten Sparzahlungen (graue Fläche) zu dem Zuzahler-Prototypen.*

kausalen Reihenfolge einen Entscheidungsbaum auf. Als Schicht wird die Kombination eines Prototypen mit einer vollständigen Folge von Bausparaktionen, die also bis zu einer Spitze des Entscheidungsbaums führt, bezeichnet. Jeder Prototyp spaltet sich über die Verzweigungen des Baumes in Schichten auf (Abb. **??**).

Die Bausparsumme, die dem Prototyp zugewiesen wurde, verteilt sich ebenfalls entlang den Verzweigungen des Baumes auf die einzelnen Schichten. Anhand derjenigen Verträge, die dem Prototypen zugeordnet wurden, werden die Anteile bestimmt, mit der sich die Bausparsumme auf die verschiedenen Verzweigungen verteilt. Auf diese Weise gelangt man zu der Bausparsumme einer einzelnen Schicht. Das vollständige Verhalten eines Bausparkollektivs wird durch die Schichten und deren Bausparsummen abgebildet. Die Simulation eines einzelnen Bausparvertrages besteht darin, dass aus den primären Größen, Sparzahlungen und Bausparaktionen, die sekundären Größen wie Guthaben, Darlehen, Zinsen, Gebühren usw. in ihrem Zeitverlauf berechnet werden. Die Simulation eines Bausparkollektives setzt sich aus den Simulationen einzelner Schichten zusammen, die unabhängig voneinander durchgeführt werden (Abb. **??**). Das Simulationsmodell ist insofern deterministisch, als mit der Bestimmung der Schichten und deren Bausparsummen das zukünftige Verhalten des Bausparkollektivs eindeutig festgelegt ist.

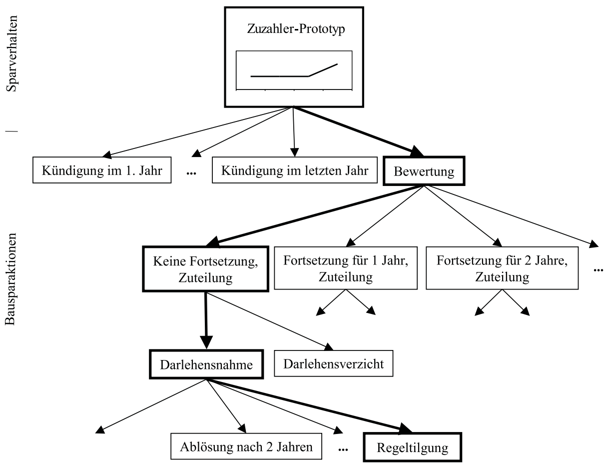

Abbildung 6.4: *Aufspaltung des Zuzahler-Prototypen in* **Schichten**. *Eine Schicht ist als Beispiel durch Fettdruck hervorgehoben.*

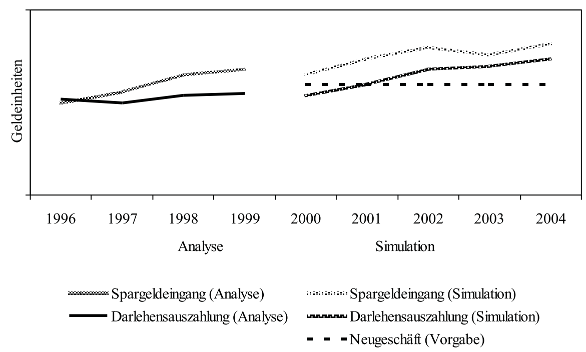

Abbildung 6.5: *Idealisierte Darstellung der* **Ergebnisse** *einer Analyse und einer anschließender Simulation unter Vorgabe eines konstanten Neugeschäfts.*

6.2 Weiterentwicklung des Modells

Inzwischen hat sich der Modellansatz bei zahlreichen Simulationen für mehrere Landesbausparkassen bewährt, jedoch gab der praktische Einsatz des Modells auch Ansatzpunkte für Verbesserungen, hauptsächlich in puncto Komplexität und Steuerbarkeit. Die Notwendigkeit der Komplexitätsreduktion im Modell ergibt sich angesichts ausgereizter Rechnerressourcen aus dem sich verändernden Bauspargeschäft, bei dem in Zukunft weitere Tarifkonstruktionen und Handlungsmöglichkeiten für die Bausparer abzubilden sind, ebenso wie aus der komplizierten Handhabbarkeit des Modells. Daneben sollte die Durchführung von Simulationen effizienter gemacht werden, als auch der seitens der Bausparkassen nötige Aufwand gesenkt werden, der mit der Bereitstellung der Kollektivdaten verbunden ist.

6.2.1 Datenbank

Zur Ermittlung der Häufigkeiten, mit denen Bausparer bestimmte Aktionen (z. B. Kündigung, Fortsetzung, Darlehensverzicht etc.) durchführen, ist eine möglichst genaue Untersuchung der Einzelvertragsdaten im Zeitverlauf notwendig. Seit diesem Jahr wird ein relationales Datenbankmanagement–System von uns eingesetzt, so dass Abfragen wesentlich schneller in SQL formuliert werden können. Durch Schnittstellen zum Cluster– und Zuordnungsprogramm können nun zudem auch relativ leicht prototypenspezifische Häufigkeiten ermittelt werden, da hierdurch in der Datenbank bekannt ist, zu welchem Prototyp ein bestimmter Vertrag zugeordnet wurde.

6.2.2 Erzeugung neuer Prototypen

Um die Anzahl der Prototypen und damit auch die der sich daraus ergebenden Schichten zu senken, wurden neue Protoypen erzeugt, wobei im Gegensatz zur ursprünglichen Clusterung nach Tarifen getrennt geclustert wurde. Diese Vorgehensweise erlaubt es, ein gutes Clusterergebnis auch mit deutlich weniger Clustern bzw. Prototypen zu erzielen, da das Sparverhalten innerhalb ein und desselben Tarifes deutlich homogener ist als über mehrere Tarife. So existieren z. B. in Tarifen mit 40% Mindestanspargrad keine 50%–Soforteinzahler und umgekehrt. Daneben wurden zusätzlich zu den Verträgen, die ihre Sparphase komplett beendet haben, auch Kündigerverträge geclustert. Diese leisten einen teilweise erheblichen Beitrag zum Spargeldeingang bis zu ihrer Kündigung, und einige Tarife bestehen fast ausschließlich aus Kündigern. Ein weiterer Vorteil der Neuclusterung besteht in der höheren Aktualität und Vollständigkeit des Datenmaterials: Da die Daten der Jahre 1997 und 1998 hinzukamen, sind die Vertragsverläufe über 14 statt über nur 12 Jahre bekannt, so dass insbesondere Langsparer besser abgebildet werden konnten.

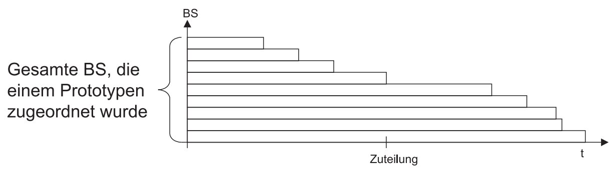

Abbildung 6.6: *Separate Berechnung einzelner Schichten im NBI.*

6.2.3 Dynamische Schichtenabspaltung und Steuerung

Das Simulationsprogramm NBI, welches die einzelnen Schichten von Vertragsanfang bis -ende durchrechnet, war ursprünglich zur Simulation einzelner Bausparverträge konzipiert. Eine Kollektivsimulation bestand aus der separaten Simulation vieler einzelner Schichten. Die große Zahl an möglichen Bausparaktionen, die einem Bausparer zur Verfügung steht, führt allerdings zu einem überproportionalen Anstieg der Schichten, in die die einem Prototyp zugeordnete Bausparsumme (in der Grafik Abb. **??** mit BS bezeichnet) zerlegt und anschließend separat durchgerechnet werden muss, so dass angesichts weitgehend ausgereizter CPU– und Speicherkapazitäten einer Berücksichtigung weiterer Wahlfreiheiten mit dem alten Ansatz enge Schranken gesetzt waren.

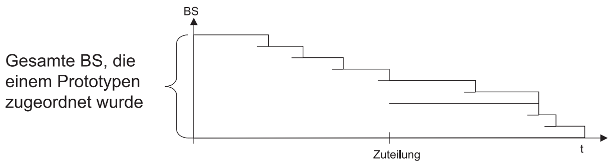

Abbildung 6.7: *Abspaltung der einzelnen Schichten im Simulationsverlauf.*

Zwar konnte das Problem des überproportionalen Anstiegs der Schichtanzahl mit steigenden Optionsmöglichkeiten nicht prinzipiell gelöst werden, allerdings ermöglicht ein geschickterer Algorithmus einen erheblichen Effizienzgewinn. Dieser beruht auf der Idee, dass eine weitere Verteilung der einem Prototypen zugeordneten Bausparsumme auf Schichten in der Simulation solange verzögert werden kann, bis das abweichende Sonderverhalten auch tatsächlich berücksichtigt werden muss (siehe Abb. **??**). Die entsprechende Schicht wird erst zu diesem Zeitpunkt von der Bausparsumme des Prototypen abgespalten. Damit wird immer mit der minimal notwendigen Anzahl Schichten gerechnet. Hinzu kommt, dass Schichten, die kündigen, auf ihr Darlehen verzichten oder das Darlehen ablösen, sofort nach ihrer Abspaltung auch wieder verschwinden, da diese zuende berechnet sind.

Damit ist wieder für Kapazitäten gesorgt, die zur Abbildung zukünftiger Tarifkonstruktionen mit weiteren Handlungsmöglichkeiten für die Bausparer nötig sind.

Im Zuge der Simulationen für die Kassen stellte sich heraus, dass die Erstellung bestimmter Szenarien die simulationszeitabhängige Variation einiger Verhaltenshäufigkeiten erfordert, um auf diese Weise z. B. Trends leichter abbilden zu können. Im Rahmen der Umsetzung der dynamischen Schichtabspaltung konnte diese Anforderung durch eine zeitabhängige Skalierung der abzuspaltenden Bausparsummen relativ leicht umgesetzt werden, so dass der künftige Aufwand zur Erzeugung bestimmter Szenarien deutlich geringer ausfallen wird.

6.2.4 Zuordnung von „kurzen Kassen“

Die Anwendung des Simulationsmodells erfordert auch von den zu simulierenden Bausparkassen einen hohen Aufwand an Vorarbeit. Da die einzelnen Verträge eines Kollektivs den Prototypen zugeordnet werden müssen, werden Daten über das Verhalten jedes einzelnen Bausparvertrages im Zeitverlauf über einen möglichst langen Zeitraum benötigt. Diese Daten zu beschaffen stellt insbesondere für Kassen, die bislang noch keine Daten geliefert haben, einen erheblichen Aufwand dar, von dem sich ein großer Teil einsparen lässt, wenn lediglich Daten über einen kurzen Zeitraum bis hin zu den Konto-Endständen des vergangenen Jahres geliefert werden können. Da eine Clusterung auf der Basis solcher „kurzer“ Daten nicht sinnvoll ist, muss auf die aus den Daten einer anderen Kasse gewonnenen Prototypen zurückgegriffen werden. Die Zuordnung solch „kurzer“ Daten zu den Prototypen ist allerdings aus dem Grund kritisch, da die Zuordnung der einzelnen Verträge durch einen Vergleich der Sparraten im Zeitverlauf mit denen der Prototypen erfolgen sollte, nämlich mit demselben Abstandsmaß, mit dem auch die Clusterung durchgeführt wurde. Da sich aber sowohl die Höhe als auch die zeitliche Verteilung der Sparraten im aktuellen Anspargrad und der Bewertungszahl eines Bausparvertrages niederschlagen, können diese Größen stellvertretend für die Sparraten zur Zuordnung verwendet werden. Hierzu werden die entsprechenden Werte für jeden Prototyp zu jedem Zeitpunkt berechnet und ein Abstandsmaß definiert, welches beide Größen so gewichtet, dass die Übereinstimmung mit der herkömmlichen Zuordnung am größten ist. Insgesamt zeigte sich bei Vergleichssimulationen mit denselben Daten die prinzipielle Eignung dieses Vorgehens.

6.3 SVD-Clusterung von Bauspar-Zeitreihen

Für die Prototypengenerierung haben wir bisher ein geometrisches Clusterverfahren, das K-means-Verfahren, verwendet, das diejenigen Verträge zu einem Cluster zusammenfasst, deren Sparverhalten im Zeitverlauf ähnlich sind. Als Ähnlichkeitsmaß haben wir den euklidischen Abstand gewählt. Der Prototyp eines Clusters wird durch den Clusterschwerpunkt definiert. Mit Hilfe dieses Verfahrens konnten wir Prototypen finden, die das Bausparkollektiv gut abbilden und auch aus der Realität bekannt sind. Alternativ zu dem K-means-Clusterverfahren haben wir eine neue Methode zur Generierung der Prototypen untersucht, die auf der Singular Value Decomposition (SVD) basiert und kein explizites Abstandsmaß verwendet. Unter der Singular Value Decomposition wird die Zerlegung einer Matrix A in das Produkt zweier orthogonaler Matrizen U und V und einer Diagonalmatrix D verstanden: $A = UDV^T$. Als Ausgangsmatrix haben wir die Bausparmatrix A so definiert, dass die Zeilen die Verträge und die Spalten die Bausparmerkmale wie Sparrate (Verhältnis zwischen Spargeldeingang und Bausparsumme), Anspargrad (Verhältnis zwischen angespartem Guthaben und Bausparsumme) oder Bewertungszahl (Kennzahl für die Leistung des Bausparers für das Bausparkollektiv) im Zeitverlauf darstellen. Die Bausparmatrix kann um weitere Merkmale wie demographische Größen (Beruf oder Alter des Bausparers) erweitert werden. Auch hier haben wir als Datenbasis alle Bausparverträge betrachtet, deren Sparphasen uns vollständig bekannt sind. Bilden wir die Singular Value Decomposition der normierten Bausparmatrix und betrachten die zu den d größten Werten der Diagonalmatrix D gehörenden Spalten der linken orthogonalen Matrix U, dann erhalten wir aus den zu den p größten positiven (negativen) Spalteneinträgen gehörenden Verträgen pro Spalte $2d$ Cluster. Die Prototypen berechnen sich als Schwerpunkte der gefundenen Cluster. Zur Berechnung der Prototypen haben wir die relationale Bauspar-Datenbank eingesetzt, mit deren Hilfe wir die Cluster und die Prototypen leicht bestimmen können. Die mittels SVD-Clusterung gefundenen Prototypen weisen einen ähnlichen Sparverlauf wie die Prototypen der K-means-Clusterung auf und können auch in der Realität wiedergefunden werden. Somit haben wir hier ein zur K-means-Clusterung alternatives Clusterverfahren an der Hand, mit dessen Hilfe wir auch weitere Merkmale außer der Sparrate zur Bestimmung der Prototypen hinzuziehen können, ohne einen expliziten Abstand zu berechnen.

6.4 Stochastische Modellierung mit Hidden–Markov–Modellen

Ein alternatives Vorgehen zur deterministischen Modellierung von Bausparkollektiven, wie es in den vorherigen Abschnitten beschrieben wurde, besteht in der stochastischen Beschreibung des möglichen Verhaltens von Bausparern. Im Rahmen zweier Dissertationen wurde hierzu ein Ansatz entwickelt und untersucht, der auf sogenannten Hidden–Markov–Modellen (HMM) basiert. Zu diesem Zweck wurde eine umfangreiche, universelle HMM–Bibliothek zur Modellierung heterogener, nicht stationärer Daten implementiert, die auch in Bereichen außerhalb des Bausparwesens verwendet werden kann. Hidden–Markov–Modelle werden seit einigen Jahren sehr erfolgreich in verschiedensten Anwendungsgebieten wie z. B. in der Bioinformatik, der Musterkennung, der automatischen Spracherkennung und der Zeitreihenanalyse eingesetzt.

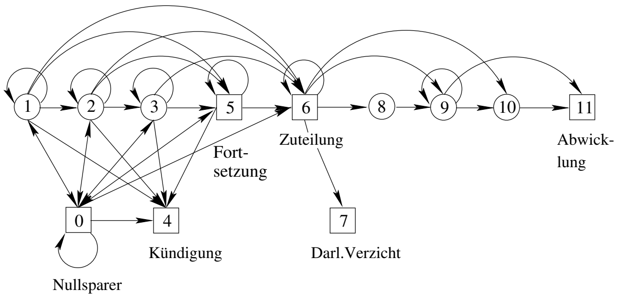

Abbildung 6.9: *Graph zur HMM–Modellierung eines Bausparvertrages.*

6.4.1 Hidden–Markov–Modelle

Ein HMM beschreibt einen stochastischen Prozess, der sich aus zwei gekoppelten Mechanismen zusammensetzt: Eine „versteckte“ Markov–Kette mit einer endlichen Anzahl von Zuständen wird in diskreten Zeitschritten durchlaufen und generiert dabei in jedem Zustand ein Ausgabesymbol gemäß einer von dem jeweiligen Zustand abhängenden Wahrscheinlichkeitsdichte. Für einen Beobachter ist nur die so entstehende Sequenz von Ausgabesymbolen sichtbar, während die darunterliegende Folge von Zuständen verborgen bleibt.

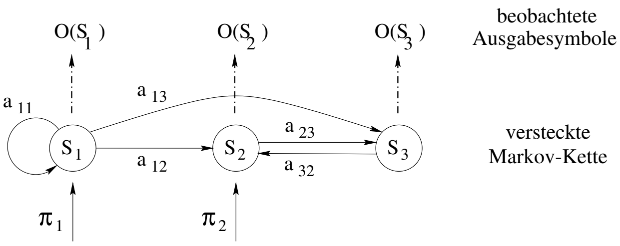

Abbildung 6.8: *Beispiel für ein einfaches HMM mit drei Zuständen.*

Abbildung **??** zeigt ein Beispiel für ein einfaches HMM mit den relevanten Modellparametern. Die Knoten des Graphen, symbolisiert durch Kreise, stehen für die möglichen Zustände, die das System einnehmen kann, und die gerichteten Kanten entsprechen jeweils den Übergängen zwischen zwei Zuständen, wobei jede Kante mit der jeweiligen Übergangswahrscheinlichkeit (a) gewichtet ist. Die zusätzlichen Eingangskanten unter den Knoten enthalten die Wahrscheinlichkeit (π), dass der Prozess in dem entsprechenden Zustand startet, und $O(S)$ kennzeichnet die Wahrscheinlichkeitsdichte, mit welcher im Zustand S die Ausgabesymbole generiert werden.

6.4.2 Modellansatz

Was hat jetzt ein solch abstraktes Modell mit dem Verhalten von Bausparern zu tun? Die Idee besteht darin, dass auch ein Bausparer während der Vertragslaufzeit verschiedene Phasen durchläuft, die zum Teil von seinen persönlichen Gegebenheiten und Intentionen abhängen und zum Teil durch äußere Faktoren festgelegt werden, die mit der Bausparkasse und dem typischen Ablauf eines Bausparvertrages zusammenhängen. Die persönlichen Umstände sind selbstverständlich schwer fassbar und in der Regel auch in den zur Verfügung stehenden Daten nicht enthalten. Dies steht im Einklang mit der Unbeobachtbarkeit der Zustände eines Hidden–Markov–Modells. Die Ausgabesymbole des Modells dagegen sind beobachtbar und entsprechen den Aktionen des Bausparers, wie z. B. Spareinzahlungen, Tilgungszahlungen, Kündigung oder Darlehensverzicht. Die zeitliche Abfolge dieser Aktionen wird im Modellansatz als eine Sequenz von Ausgabesymbolen aufgefasst, und das gesamte Bausparkollektiv besteht in dieser Modellierung folglich aus einer entsprechend großen Anzahl einzelner Sequenzen.

In Abbildung **??** ist die Topologie des in der Bausparsimulation eingesetzten Gesamtmodells dargestellt. Zustände, die mit einem Quadrat gekennzeichnet sind, spielen eine besondere Rolle, da sie lediglich ein fest definiertes, den jeweiligen Zustand kennzeichnendes Symbol ausgeben können. Hierdurch ist es möglich, Zustände fest mit bestimmten Aktionen zu verknüpfen. Außerdem wird durch die Einschränkung der Übergänge der prinzipielle Ablauf eines Bausparvertrages vorgegeben.

Ein Grund für den verbreiteten Einsatz von Hidden–Markov–Modellen ist darin zu sehen, dass mit dem Baum–Welch–Algorithmus ein effektives Verfahren vorhanden ist, das ein Training der Modelle und damit eine Anpassung an vorhandene Daten ermöglicht. Die vom Algorithmus ermittelten Modellparameter stellen ein lokales Optimum im Parameterraum dar. Sie werden also durch das Training so ermittelt, dass die vorgegebenen Daten möglichst gut vom Modell repräsentiert werden.

6.4.3 Training und Erweiterung der Hidden–Markov–Modelle

Das Training der Hidden–Markov–Modelle geschieht mit Bausparverträgen, deren Spar– oder Darlehensphase vollständig abgeschlossen ist. Dabei werden mehrere, zunächst recht unspezifische Modelle vorgegeben, die sich im Laufe des Trainings auf eine Teilmenge der Verträge spezialisieren können. Dies stellt eine Analogie zur in den vorhergehenden Abschnitten beschriebenen Prototypenbildung durch Clusterung dar, wobei diese Teilmengen den Clustern entsprechen und die Parameter der trainierten Hidden–Markov–Modelle den Prototypen. Im deterministischen Bauspar–Modell, wie in den vorhergehenden Abschnitten beschrieben, ergeben sich die Prototypen durch eine Mittelung über die Verträge im Cluster. Im Fall der Hidden–Markov–Modelle werden die statistischen Eigenschaften der Sequenzen durch die Modellparameter abgebildet, die entstehenden Prototypen sind also stochastischer Natur. Um diese stochastischen Prototypen zu gewinnen, wird das Training der Modelle durch den Baum–Welch–Algorithmus mit zwei bekannten Clusterverfahren, dem K–Means–Verfahren und dem EM–Algorithmus zur Clusterung mit Mischmodellen, kombiniert.

Um Hidden–Markov–Modelle erfolgreich zur Simulation von Bauspardaten einzusetzen, wurden die Standardmodelle in einigen Punkten erweitert. Dies umfasst unter anderem die Integration von deterministischen Nebenbedingungen, die aus gesetzlichen und tariflichen Vorgaben resultieren. Desweiteren ist es notwendig, Sequenzen mit einer Gewichtung zu versehen, da in der Realität Verträge immer mit einer Bausparsumme verbunden sind. Außerdem muss durch die Wahl einer geeigneten Wahrscheinlichkeitsdichte für die Ausgaben sichergestellt werden, dass keine negativen Spar- bzw. Tilgungszahlungen vom Modell erzeugt werden, da dies in der Realität nicht vorkommt.

6.4.4 Simulationen

Sind die optimalen Modellparameter gefunden, so ist es über einen Monto–Carlo–Ansatz sehr einfach möglich,

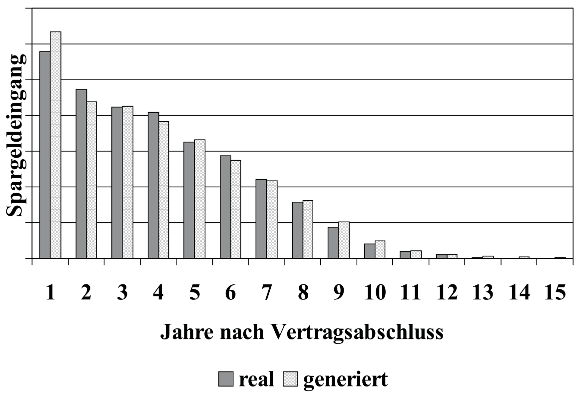

Abbildung 6.10: *Verteilung der Spargeldeingänge bei realen Trainingsdaten und bei künstlichen, mit einem HMM generierten Sequenzen.*

neue, künstliche Sequenzen zu generieren oder auch bestehende Teilsequenzen zu vervollständigen. Hierzu werden unter Verwendung eines Zufallszahlengenerators basierend auf den Modellparametern neue Markov–Ketten und dazugehörige Ausgabesymbole generiert. Die Anteile der verschiedenen Modelle an den künstlichen Sequenzen ergeben sich unmittelbar aus den jeweiligen Volumina der Modelle in der Clusterung.

Zur Verlängerung vorhandener Teilsequenzen muss zunächst bestimmt werden, welchem der zur Auswahl stehenden Hidden–Markov–Modelle die Sequenz entspringt. Da dies in der Regel nicht eindeutig möglich ist, wird hierfür das Modell mit der größten Wahrscheinlichkeit zur Erzeugung dieser Sequenz gewählt, wobei es über einen Bayes–Ansatz möglich ist, a priori Wahrscheinlichkeiten der verschiedenen Modelle zu berücksichtigen. Dies hat Relevanz bei einer evtl. eingreifenden Steuerung von Simulationen zur Durchführung von Szenarien und bei der Verlängerung sehr kurzer Teilsequenzen, bei denen die Zuordnung zu einem bestimmten Modell zwangsläufig mit einer gewissen Unsicherheit verbunden ist. Ist das Generatormodell bestimmt, so wird im nächsten Schritt der wahrscheinlichste Zustand dieses Modells zum Zeitpunkt des Sequenzendes berechnet, um hiervon ausgehend neue Symbole zu generieren.

6.4.5 Ergebnisse

Diverse Untersuchungen mit dem beschriebenen Modellansatz haben gezeigt, dass Hidden–Markov–Modelle sehr gut geeignet sind, die stochastischen Charakteristika der

Bauspar–Zeitreihen zu modellieren. Beispielhaft zeigt Abbildung **??** einen Vergleich der Spargeldaufkommen in den verschiedenen Jahren nach Vertragsabschluss von realen Trainingsdaten und einer künstlich generierten Menge von Sequenzen.

Die beschriebenen Erweiterungen führen zu einer deutlich verbesserten Abbildung der relevanten Größen gegenüber dem Standard–HMM. Erste Vergleiche eines generierten Gesamtkollektivs mit den entsprechenden realen Daten sind vielversprechend und lassen einen erfolgreichen Einsatz des Modells zur Durchführung von Simulationen in der Praxis erwarten.

6.5 Ausblick

Die Weiterentwicklung des Simulationsmodells hängt stark von den Zielsetzungen der LBS ab. Als wichtiger Punkt ist sicherlich die Implementierung eines flexibilisierten Bewertungs– und Zuteilungsverfahrens zu nennen. Ein solches Verfahren wird für das Simulationsmodell neben eher technischen Fragen der konkreten Realisierung insbesondere auch grundsätzliche Probleme der Bestimmung bestimmter Parameter aufwerfen. Hier stellt sich die Frage, wie bei neu einzuführenden Handlungsmöglichkeiten für die Bausparer die entsprechenden Häufigkeiten abgeleitet werden können. Interessant wären auch tiefergehende Untersuchungen im Zusammenhang mit der dynamischen Steuerung der Simulationen und hier insbesondere der Einfluss äußerer Faktoren auf das Kollektivverhalten. Mit der beschriebenen dynamischen Schichtenabspaltung wurde die Voraussetzung einer effektiven Steuerung geschaffen.

Im Rahmen der stochastischen Modellierung mit Hidden–Markov–Modellen ergeben sich sowohl aus theoretischer als auch aus praxisorientierter Sicht eine Reihe weiterer interessanter Fragestellungen und Aufgaben, die den Gegenstand einer wissenschaftlichen Arbeit bilden können. Im theoretischen Bereich ist sicherlich die Entwicklung von Verfahren zur Überwindung lokaler Maxima beim Training der Modelle und bei der Clusterung ein lohnenswertes Projekt. Einer Anwendung dieses Modells in der Praxis müsste auf jeden Fall die Kopplung mit dem bestehenden Rechenkern NBI zur Berechnung verschiedener sekundärer Größen wie Zinsen und Gebühren und zur Durchführung von Szenariorechnungen vorausgehen.

Kontakt: `bauspar@zpr.uni-koeln.de`