

Department for Applied Statistics
Johannes Kepler University Linz



IFAS Research Paper Series 2007-27

Methoden zur Quotenprognose und statistischen Risikoanalyse für die Auszahlungen des Allgemeinen Entschädigungsfonds für die Opfer des Nationalsozialismus

Helmut Waldl, Andreas Quatember und Werner G.
Müller

July 2007

1 Zusammenfassung

Ende Dezember 2005 konnte mit den ersten Auszahlungen des Allgemeinen Entschädigungsfonds (General Settlement Fund, kurz GSF) begonnen werden. Im entscheidenden Gutachten zu den Auszahlungssummen konnten unter Verwendung einer sogenannten Imputationsmethode die noch fehlenden Auszahlungssummen geschätzt und auf deren Basis Vorauszahlungsquoten festgelegt werden. Die Vorauszahlungen betragen je nach Kategorie zehn oder 15 Prozent der als berechtigt anerkannten Forderungen. Zehn Prozent gibt es für Ansprüche aus dem sogenannten „Forderungsverfahren“. Dieses kommt zur Anwendung, wenn die Ansprüche gut belegt und nachweisbar sind. Das „Billigkeitsverfahren“ mit einer Vorauszahlungsquote von 15 Prozent greift in erster Linie bei ideellen Verlusten, etwa wegen einer abgebrochenen Berufsausbildung oder dem Verlust von Berufsrechten. Ebenfalls 15 Prozent beträgt die Quote für die Entschädigung nicht erstatteter Versicherungspolizzen.

2 Juristische und empirische Grundlagen

Der Allgemeine Entschädigungsfonds für Opfer des Nationalsozialismus soll vorläufige Leistungen an seine Antragsteller erbringen. Dazu wird eine Prognose der Auszahlungsquoten verbunden mit einer Risikoanalyse als Basis für die Wahl geeigneter Auszahlungsquoten erstellt. Ziel dieser Analyse war es, dem Kuratorium des Fonds die Festlegung einer Vorauszahlungsquote unter Abwägung zwischen dem Ziel der Vermeidung zu hoher vorläufiger Leistungen und dem Ziel der Erbringung möglichst hoher vorläufiger Leistungen an die Antragsteller zu ermöglichen.

Für das Billigkeitsverfahren, das Forderungsverfahren und das Verfahren für Forderungen aus entzogenen Versicherungen sollten abhängig von den Vorauszahlungsquoten Wahrscheinlichkeiten für ein Überschreiten der endgültigen Auszahlungsquoten bestimmt werden.

Die juristischen Grundlagen für die Vorauszahlungen sind einerseits das Entschädigungsfondsgesetz (BGBl. I Nr.12/2001) vom 28.2.2001 und im Besonderen die Abänderung des Entschädigungsfondsgesetzes (BGBl. I Nr.142/ 2005) vom 13.12.2005.

Konkrete Daten dürfen aufgrund der vertraglich vereinbarten Verschwiegenheitspflicht erst veröffentlicht werden, wenn der Fonds seine Arbeiten abgeschlossen hat. Bei den Daten, die in dieser Arbeit ersichtlich werden, handelt es sich ausschließlich um solche, die bereits veröffentlicht wurden. Aber auch ohne konkrete Daten sind die theoretischen Grundlagen der Methodik und die bei der praktischen Umsetzung auftauchenden Probleme von Interesse.

Basis der Analyse waren eine vom Fonds zur Verfügung gestellte MS-Access Datenbank und spätere korrigierte Versionen dieser Datenbank.

3 Theoretische Grundlagen der gewählten Vorgehensweise

Bei der statistischen Datenanalyse ist das Auftreten von fehlenden Werten ein bekanntes und selbst bei den sorgfältigst durchgeführten Erhebungen auftretendes Fak-

tum. Spezielle Methoden wie die Imputationsmethoden wurden entwickelt, die auf diese Problematik eingehen (für einen Überblick zur Problematik siehe etwa: Särndal, Swensson, and Wretman (1992), Kapitel 14-16). Bei unserer Aufgabe liegt eine Grundgesamtheit U („universe“) an zu bearbeitenden Forderungen vor. Diese Menge U an Forderungen lässt sich zerlegen in eine Teilmenge R („response“) an Fällen, für die das interessierende Merkmal y , die Forderungshöhe, bereits bestimmt wurde, und eine Teilmenge M („missing“) an Fällen, die noch nicht bearbeitet wurden ($U = R + M$).

Gesucht wird die Merkmalssumme t von y :

$$t = \sum_U y_k = \sum_R y_k + \sum_M y_k$$

Der zweite Summand auf der rechten Seite bezieht sich auf die Menge M der Fälle mit fehlenden Werten. Zur Schätzung dieser Summe werden bei den Imputationsmethoden unter Verwendung von zu den einzelnen Fällen vorhandenen Hilfsinformationen Ersatzwerte y_k^i ($i \dots$ imputiert) für y_k ($k \in M$) gesucht. Der Schätzer \hat{t} für t ist dann:

$$\hat{t} = \sum_R y_k + \sum_M y_k^i$$

Zerlegt man die Grundgesamtheit U in K Schichten h ($h = 1, \dots, K$), so lässt sich \hat{t} durch Summation über alle Schichten darstellen:

$$\hat{t} = \sum_h \left(\sum_{R_h} y_k + \sum_{M_h} y_k^i \right)$$

mit U_h , der Menge der Fälle in der h -ten Schicht, die sich wiederum in eine Responsemenge R_h und eine Missingmenge M_h zerlegen lässt.

Die gewählte Imputationsmethode ist *hot deck within classes*, wobei innerhalb der gebildeten Schichten aus den in R beobachteten Werten die Ersatzwerte für M zufällig mit Zurücklegen gezogen werden. Diese Single Imputations-Methode ermöglicht jedoch keine korrekte direkte Schätzung der Streuung des Schätzers \hat{t} , da ein mit den Ersatzwerten y_k^i auf herkömmliche Weise berechneter Varianzschätzer nur den Stichproben-, nicht jedoch den Imputationsfehler miteinbezieht und die tatsächliche Streuung, also die Ungenauigkeit der Schätzung, somit unterschätzt. Im vorliegenden Sonderfall, in dem die fehlenden Werte in U selbst und nicht in einer Stichprobe aus U auftreten, würde die auf diese Weise errechnete Schätzung der Varianz sogar null sein.

Die *multiple Imputation* von gleich D Ersatzwerten pro fehlendem Wert hingegen ermöglicht eine Varianzschätzung, die den Imputationsfehler als Streuungsquelle mitberücksichtigt, indem durch die Erzeugung von D Datensätzen zum Stichprobenfehler noch die rein durch die Imputationen verursachte Streuung zwischen den daraus resultierenden D Schätzern \hat{t} addiert werden kann (siehe dazu: Rubin (1987)). Da bei Multipler Imputation von Ersatzwerten mittels des Verfahrens Hot Deck within classes jedoch die Zufälligkeit der Responsegruppe immer noch nicht mitberücksichtigt wird (und somit die tatsächliche Variabilität des Schätzers trotz multipler Imputation weiterhin unterschätzt werden würde), wird die Methode noch modifiziert. Dabei werden nach den mittels *hot deck within classes* erfolgten Imputationen in jedem der D Imputationsschritte schichtweise erst durch abermalige Ziehung mit Zurücklegen aus der so gewonnenen prädiktiven Dichte des Merkmals y in M die

tatsächlich verwendeten Ersatzwerte y_k^i bestimmt. Der erste Schritt spiegelt nun die Unsicherheit der theoretischen Verteilung von y , aus der gezogen werden soll, wider. Dieses Verfahren wird als *Approximate Bayesian Bootstrap* bezeichnet (siehe: Little and Rubin (2002), Seite 85ff), wobei jedoch jeweils nur eine Bootstrap-Stichprobe gezogen wird. Rubin and Schenker (1986) zeigen, dass für große Anzahlen D an Multiplen Imputationen auf diese Weise unter bestimmten Voraussetzungen schichtweise unverzerrte Schätzungen der Verteilung von \hat{t} erzeugt werden und sich die Verteilung der D Schätzwerte somit in diesen Fällen für die Angabe von Konfidenzbereichen für den gesuchten Parameter t eignet.

4 Datenaufbereitung

Ein wesentlicher Teil der Arbeiten des Projektes bestand in der Vorbereitung der eigentlichen Analyse: Die Daten mussten aufbereitet und eventuelle Zusammenhänge zwischen den vorhandenen Variablen untersucht werden.

Die Untersuchung möglicher Zusammenhänge zwischen den Variablen erfolgte mit dem Statistikprogrammpaket SPSS. Dazu mussten aus den vorhandenen Variablen einige neue Variablen berechnet werden, dies geschah zum Teil durch SQL-Abfragen in der Datenbank selbst oder im Programmpaket SPSS. Bei der Analyse mit SPSS wurden die Daten nicht direkt aus der Datenbank eingelesen, sie mussten aus der Datenbank in einem geeigneten Format exportiert und von SPSS importiert werden.

Zur Analyse der Zusammenhänge zwischen den Variablen wurden Kontingenztafeln erzeugt und Chi-Quadrat-Tests durchgeführt. Bei dieser ersten Datenanalyse tauchten auch teilweise unerklärliche Ergebnisse auf, die schließlich zum Entdecken einer fehlerhaften Variablen in der ursprünglichen Version der Datenbank beitrugen. Mit den Daten der neuen Version der Datenbank musste die Analyse möglicher Zusammenhänge von Variablen mit der korrigierten Variablen wiederholt werden. Aber auch hier tauchten Ungereimtheiten auf, die ihre Ursache darin hatten, dass sich für die Tabellen der neuen Datenbank die Primärschlüssel im Vergleich zur alten Datenbank geändert hatten. Nach Abgleichung dieser Primärschlüssel in beiden Datenbanken musste also die Analyse mit der korrigierten Variablen erneut durchgeführt werden. Die Ergebnisse dieser ersten Analyse waren schließlich erfolgversprechend, es wurden einige wichtige Zusammenhänge zwischen den Variablen der Datenbank bzw. zwischen neu berechneten Variablen entdeckt.

Nun konnte die eigentliche Analyse, die Quotenschätzung beginnen. Dieser Teil der Analyse wurde mit MATLAB, einem Hochleistungsprogrammpaket für numerische Berechnungen durchgeführt. Dazu mussten die Daten teils aus SPSS teils aus der Datenbank in einem für MATLAB geeigneten Format exportiert und von MATLAB eingelesen werden. Da es sich bei dem angewandten Verfahren um ein sehr rechenintensives handelt, musste nach der eigentlichen Programmierung der Routinen noch Arbeit in die Verbesserung der benötigten Rechenzeit investiert werden. Und schließlich wurde noch einige Zeit in die Programmverifikation gesteckt.

5 Beschreibung der Analyse

Da die Schätzer für die Forderungssummen (die Summe aller bewerteten Forderungen, nachdem alle Akten bearbeitet sind) in den drei verschiedenen Verfahrensarten jeweils Summen und Produkte von untereinander zum Teil abhängigen Zufallsvariablen sind, ist es praktisch unmöglich, die Variabilität der Schätzer für die Forderungssummen auf theoretischem Wege zu bestimmen. Wir haben uns daher entschieden, die Verteilungen der Schätzer für die Forderungssummen und anderer Variablen mit Hilfe *multipler Imputation* zu schätzen. Es wurde sozusagen die gesamte Verteilung der Schätzer für die Forderungssummen etc. erzeugt, und zwar folgendermaßen:

Die fehlenden Daten (die Daten der nicht bearbeiteten Antragsteller und der noch nicht bearbeiteten Forderungen) wurden durch *Hot Deck Imputation* ergänzt. Das heißt, für fehlende Daten wurden vorhandene Daten von Antragstellern bzw. Forderungen mit ähnlicher Struktur eingesetzt.

Beispiel: bei der Forderung eines österreichischen Antragstellers mit Forderungskategorie Betriebe, die sich auf eine weibliche historische Person bezieht, fehlt die Forderungssumme. Man zieht daher zufällig eine Forderungssumme aus den vollständig vorhandenen Datensätzen, bei denen der Antragsteller „Österreicher“, die Kategorie „Betriebe“ und die historische Person „weiblich“ ist.

Wie wir dabei genau vorgegangen sind, wird unten detailliert beschrieben. Hat man das löchrige Datenmaterial auf diese Art vervollständigt, kann man die Schätzer für die Forderungssummen oder beliebige andere Statistiken genauso wie bei vollständigen Daten berechnen. Das Problem der so berechneten Schätzer für die Forderungssummen ist, dass sie Zufallsschwankungen unterworfen sind. Deshalb führt man bei *multipler Imputation* dieses Auffüllen der fehlenden Daten und das Berechnen der Schätzer für die Forderungssummen mehrmals - möglichst oft - durch und erhält auf diese Weise die Wahrscheinlichkeitsverteilung der Schätzer für die Forderungssummen. Kennt man diese Wahrscheinlichkeitsverteilung, kann man Bereichsschätzer bestimmen, also Intervalle, die die unbekanntes aber festen Forderungssummen mit vorgegebener Wahrscheinlichkeit überdecken. Da diese wiederholte Datenimputation sehr rechenintensiv ist, kann man sie nicht beliebig oft durchführen. Wir entschieden uns für eine Wiederholungszahl von 2000, die dazu benötigte Rechenzeit war mit etwa zehn Stunden bereits hoch genug.

Die Ergebnisse und insbesondere die Korrektheit einer Analyse sind einerseits von der Korrektheit der Daten, immer aber auch von den verwendeten Modellannahmen abhängig. Stimmen die Modellannahmen nicht, dann sind die Ergebnisse einer Analyse meist unbrauchbar. In der Praxis wird solchen Annahmen oft nicht die ihnen gebührende Bedeutung geschenkt. Wir möchten die Wichtigkeit solcher Annahmen aber besonders unterstreichen, und kennzeichnen die getroffenen Annahmen daher als solche.

5.1 Nicht erfasste Antragsteller imputieren

Der Bearbeitungsstand eines Antragstellers kann im Wesentlichen in die drei folgenden Kategorien eingeteilt werden:

- nicht erfasste Antragsteller: Forderungen der Antragstellers sind unbekannt.

- erfasste Antragsteller: Forderungen der Antragsteller sind bekannt aber nicht bearbeitet.
- bearbeitete Antragsteller: Forderungen der Antragsteller sind bearbeitet und bewertet.

Das Auffüllen der fehlenden Daten erfolgte in mehreren Schritten. Zuerst musste für 6207 nicht erfasste Antragsteller eine Forderungsstruktur (die Menge aller Forderungen eines Antragstellers) imputiert werden.

Die Forderungsstruktur wurde bei der ersten Quotenprognose aus 12967 Antragstellern gezogen, die entweder erfasst waren oder bereits eine Forderungsstruktur besaßen. Von diesen 12967 Antragstellern waren 7127 nicht bearbeitet.

Eine geringe Anzahl von Antragstellern, die erfasst waren, aber keine zugehörige historische Person und keine Forderung hatten, wurde gelöscht. Ebenso wurde bei wenigen Antragstellern jeweils eine historische Person gelöscht, zu der keine Forderungen existierten.

Bei der Imputation der Forderungsstrukturen wurde die Tatsache berücksichtigt, dass unter den bearbeiteten Antragstellern jene mit geringer Forderungszahl und jene mit Forderungen der Kategorie berufs- und ausbildungsbezogene Verluste stark überrepräsentiert waren.

Modellannahme 1: Es wurde davon ausgegangen, dass die bereits erfassten aber nicht bearbeiteten Antragsteller bezüglich der Verteilung der Forderungszahl und der Anzahl von Forderungen der Kategorie berufs- und ausbildungsbezogene Verluste repräsentativ für die nicht erfassten Antragsteller sind.

Bei der Imputation wurden die bearbeiteten Antragsteller daher entsprechend gewichtet. Für die imputierten Antragsteller wurden bei allen Forderungen die Werte der folgenden Variablen übernommen:

Vermögenskategorie des Verlustes
 die vom Antragsteller gewünschte Verfahrensart
 Ausweitung
 Nacherfasst
 das Geschlecht der historischen Person

Modellannahme 2: Es wurde davon ausgegangen, dass die gewichtete Menge der Antragsteller, aus der imputiert wurde, bezüglich der Forderungsstruktur und der obigen Variablen repräsentativ für die nicht erfassten Antragsteller ist.

Nicht übernommen wurde der Wohnsitzstaat des Antragstellers (er ist auch bei nicht erfassten Antragstellern bekannt) und die Bewertung der Forderung. Außerdem wurde unterschieden, ob aus bearbeiteten oder aus nur erfassten Antragstellern imputiert wurde. Bei den Imputationen aus den bearbeiteten Antragstellern wurden auch die Variablen {Verfahrensart, in der die Forderung entschieden wurde; Erbquote; Summe der Erbquoten der zugehörigen historischen Person} übernommen. Außerdem wurden auch die verworfenen Forderungen dieser Antragsteller übernommen, ohne sie aber als solche zu kennzeichnen.

In diesem Stadium der Analyse hat jeder Antragsteller eine (vorläufige) Forderungsstruktur. Bei den nicht bearbeiteten Antragstellern können jetzt noch Forde-

rungen durch Ausweitung und Nacherfassung dazukommen bzw. durch Verwerfen verschwinden.

5.2 Nacherfassungen und Ausweitungen simulieren

5.2.1 Nacherfassungen

Vergleicht man den Anteil der nacherfassten Forderungen an allen Forderungen der bearbeiteten Antragsteller mit jenen der nur erfassten Antragsteller, dann scheint dieser Anteil während der juristischen Bearbeitung geringfügig zu steigen. Berücksichtigt man den Anteil der nacherfassten Forderungen unter den nur erfassten Antragstellern, dann ergibt sich daraus, dass die Anzahl der nacherfassten Forderungen bei der juristischen Bearbeitung um einen bestimmten Prozentsatz wächst.

Betrachtet man die Anteile der nacherfassten Forderungen gemessen an allen Forderungen unter den bearbeiteten bzw. nur erfassten Antragstellern gesondert nach der Vermögenskategorie des Verlustes, dann ist bei einigen Kategorien ein sonderbarer Effekt zu beobachten: Besonders bei den Kategorien „Immobilie“ und „Schuldverschreibungen“ war unter den Forderungen der bearbeiteten Antragsteller der Anteil der nacherfassten Forderungen deutlich geringer als unter den Forderungen der nur erfassten Antragsteller. Vergleicht man die bearbeiteten Forderungen (Forderung ist verworfen, zuerkannt oder abgelehnt) mit den nicht bearbeiteten, dann ist der Unterschied der obigen Anteile sogar noch größer.

Für einige Vermögenskategorien galt also: Der größere Anteil der nacherfassten Forderungen befindet sich unter den noch nicht bearbeiteten Forderungen. Bedenkt man, dass während der Bearbeitung der Akte noch Forderungen durch Nacherfassung hinzukommen können, dann bedeutet dies, dass bei den anfangs bearbeiteten Akten nicht so häufig nacherfasst wurde. Vergleicht man die mittlere Anzahl von Forderungen der bearbeiteten Antragsteller mit der eindeutig größeren der nur erfassten Antragsteller, dann spricht auch das für diese Begründung. Die Anzahl der nacherfassten Forderungen muss bei der Bearbeitung der noch nicht entschiedenen Forderungen also um mehr als den oben errechneten Prozentsatz steigen.

5.2.2 Ausweitungen

Hier war die Situation ähnlich wie bei den nacherfassten Forderungen, ja sogar etwas „schlimmer“. Die Anteile der Ausweitungen waren unter den Forderungen der bearbeiteten Antragsteller und unter jenen der nur erfassten Antragsteller etwa gleich. Besonders in den Kategorien Immobilien, sonstige Verluste und Versicherungen war aber der Anteil der Ausweitungen bei den nur erfassten Antragstellern zum Teil deutlich größer als bei den bearbeiteten Antragstellern. Das bedeutet, dass auch Ausweitungen bei den schon früher bearbeiteten Antragstellern seltener vorkamen.

Eine mögliche Erklärung liefert die Tatsache, dass unter den bearbeiteten Antragstellern eine überproportionale Anzahl an Akten mit ausschließlich berufs- und ausbildungsbezogenen Verlusten und an Akten mit geringer Forderungszahl ist. Diese Akten sind einfach strukturiert, daher musste man seltener nacherfassen oder ausweiten.

Modellannahme 3: Es wurde davon ausgegangen, dass die bereits erfassten aber nicht bearbeiteten Antragsteller bezüglich der Ausweitungen und Nacherfassungen repräsentativ für die nicht erfassten Antragsteller sind.

Da diese Annahme bereits bei der Imputation der Forderungen für nicht erfasste Antragsteller berücksichtigt wurde, stellt die ungleiche Verteilung der Ausweitungen und Nacherfassungen unter den nur erfassten bzw. unter den fertig bearbeiteten Antragstellern kein Problem dar.

Es bleibt das Problem mit den zusätzlichen Ausweitungen und Nacherfassungen während der juristischen Bearbeitung. Nach den Angaben des Projektkoordinators des Allgemeinen Entschädigungsfonds für Opfer des Nationalsozialismus, Mag. Hubinger erfolgen „... Ausweitungen bei der juristischen Bearbeitung nur sporadisch ...“ und wird nur „... ein kleiner Teil der nacherfassten Forderungen von den Juristen erzeugt ...“. Wie groß dieser Anteil genau ist, konnte nicht exakt quantifiziert werden.

Modellannahme 4: Es wurde davon ausgegangen, dass sich während der Bearbeitung der Akte die Anzahl der Ausweitungen und die Anzahl der Nacherfassungen um Prozentsätze erhöhen, die aufgrund der beobachteten Daten plausibel erschienen. Da diese Annahmen eher vage waren, waren die obigen Prozentsätze Programmparameter, die für neue Programmdurchläufe jederzeit geändert werden konnten.

Im Programm wurde die Vermehrung der Ausweitungen und Nacherfassungen realisiert, indem die oben festgesetzten Anteile an Ausweitungen bzw. Nacherfassungen der nicht bearbeiteten Antragsteller zufällig ausgewählt und anschließend dupliziert wurden.

5.3 Verwerfungen simulieren

Bei den verworfenen Forderungen waren ganz klare Abhängigkeiten zu beobachten: zum Beispiel wurden von den nacherfassten Forderungen weniger verworfen als vom Rest, oder von den Ausweitungen mehr als vom Rest. Der Anteil der verworfenen Forderungen war auch stark von der Vermögenskategorie des Verlustes abhängig. Das Verwerfen findet zum Teil schon bei der Erfassung der Akten statt, hauptsächlich aber bei der juristischen Bearbeitung.

Modellannahme 5: Es wurde davon ausgegangen, dass sich der Prozess des Verwerfens im Zeitverlauf nicht geändert hat (dass also nicht etwa zu einem späteren Zeitpunkt mehr oder weniger verworfen wurde als früher) und dass die Forderungen der bearbeiteten Antragsteller die Anteile der verworfenen Forderungen nach der juristischen Bearbeitung für die „zukünftige“ Aktenerledigung gut repräsentierten.

Bei den nicht entschiedenen und den imputierten Forderungen wurde entsprechend den Anteilen der verworfenen Forderungen in den fertig bearbeiteten Akten, geschichtet nach {Vermögenskategorie des Verlustes; Ausweitung; Nacherfasst}, zufällig verworfen.

5.4 Änderung der Verfahrensart simulieren

Nach dem Verwerfen von Forderungen musste für alle nicht entschiedenen Forderungen außer jenen, die aus bearbeiteten Antragstellern imputiert wurden, die Verfahrensart, in der über die Forderung entschieden wird, simuliert werden. Bei der Datenanalyse konnte unter den entschiedenen Forderungen eine klare Abhängigkeit der Verfahrensart, in der die Forderung entschieden wurde, von der Vermögenskategorie des Verlustes, von der Variable Ausweitung und der vom Antragsteller gewünschten Verfahrensart festgestellt werden. So wechselte zum Beispiel die Verfahrensart bei Ausweitungen kaum – verständlich, wenn man bedenkt, dass hier die vom Antragsteller gewünschte Verfahrensart ja vom Bearbeiter des Aktes bestimmt wird.

Modellannahme 6: Es wurde davon ausgegangen, dass der Wechsel der Verfahrensart unter den entschiedenen Forderungen repräsentativ für alle Forderungen ist.

Die Verfahrensart, in der die Forderung entschieden wird, wurde entsprechend den unter den entschiedenen Forderungen gefundenen Anteilswerten dieser Variablen, geschichtet nach {Ausweitung; Vermögenskategorie des Verlustes; vom Antragsteller gewünschte Verfahrensart} simuliert. Bei den Forderungen, die aus bearbeiteten Antragstellern imputiert wurden, war das nicht notwendig, für sie war die Verfahrensart, in der die Forderung entschieden wurde, bekannt.

5.5 Imputation der Bewertung der Forderungen

Jetzt konnte für die nicht bearbeiteten und imputierten Forderungen die Bewertung der Forderungen imputiert werden, und zwar wurde aus abgelehnten und zuerkannten Forderungen imputiert. Da sich herausstellte, dass die Höhe der bewerteten Forderung mit den Variablen Vermögenskategorie des Verlustes, Verfahrensart, in der die Forderung entschieden wird, Wohnsitzstaat des Antragstellers und Geschlecht der historischen Person zusammenhängt, wurde die Menge der abgelehnten und zuerkannten Forderungen nach diesen Variablen geschichtet, und zwar so, dass in jeder Schicht eine Mindestzahl von 100 Beobachtungen vorhanden war. Bei den meisten Schichten war die Anzahl von Beobachtungen aber wesentlich größer. Insgesamt wurde aus 80 verschiedenen Schichten imputiert, und zwar nach dem *Approximate Bayesian Bootstrap* - Verfahren. Die verschiedenen Ausprägungen der Variablen, die zur Schichtung verwendet wurden sind der folgenden Aufstellung zu entnehmen:

- Vermögenskategorie des Verlustes:
 - Aktien, Wertpapiere
 - Bankguthaben
 - Beruf, Ausbildung
 - Betrieb
 - bewegliches Vermögen
 - Hypothekarforderung
 - Immobilien
 - Schuldverschreibungen
 - sonstige Verluste

- Versicherungen
- Verfahrensart, in der die Forderung entschieden wurde:
 - Billigkeitsverfahren
 - Forderungsverfahren
- Wohnsitzstaat des Antragstellers:
 - Australien, Neuseeland
 - Israel
 - Österreich
 - USA, Kanada
 - restliche Staaten
- Geschlecht der historischen Person:
 - männlich
 - neutral (Vereine, Organisationen)
 - weiblich

Modellannahme 7: Es wurde davon ausgegangen, dass die Bewertungen der Forderungen in den Schichten der entschiedenen Forderungen repräsentativ für alle Forderungen sind.

Zusätzlich werden für alle nicht aus bearbeiteten Antragstellern imputierten Forderungen auch noch die Variablen Erbquote und Summe aller Erbquoten der historischen Person imputiert.

5.6 Miterben simulieren

Miterben können bei historischen Personen vorkommen, wenn die Summe der Erbquoten aller Antragsteller mit dieser historischen Person < 1 ist. Unter den entschiedenen Antragstellern war dies bei $x\%$ der Fall. Allerdings war zu bedenken, dass dieser Anteil unter Berücksichtigung noch nicht bearbeiteter Antragsteller sicher etwas kleiner sein wird.

Modellannahme 8: Es wurde davon ausgegangen, dass der obige Anteil der Antragsteller mit potentiellen Miterben repräsentativ ist für alle Antragsteller, insbesondere aber auch für jene, die bereits eine Entscheidung zugesandt bekamen.

Modellannahme 9: Es wurde davon ausgegangen, dass der Anteil der tatsächlich ins Verfahren eingebrachten Miterben unter den Antragstellern mit potentiellen Miterben $x\%$ ist.

Da kein Datenmaterial für eine genauere Schätzung der Anzahl der ins Verfahren kommenden Miterben vorhanden war, machten wir den Anteil der tatsächlichen Miterben unter den Antragstellern, welche Miterben einbringen können, ebenfalls zu einem für einen weiteren Programmdurchlauf frei wählbaren Parameter.

Die Miterben wurden simuliert, indem $x\%$ der Forderungen mit einer historischen Person, deren Summe der Erbquoten < 1 war, zufällig auf Forderungen mit der maximal möglichen Erbquote von 1 ergänzt wurden.

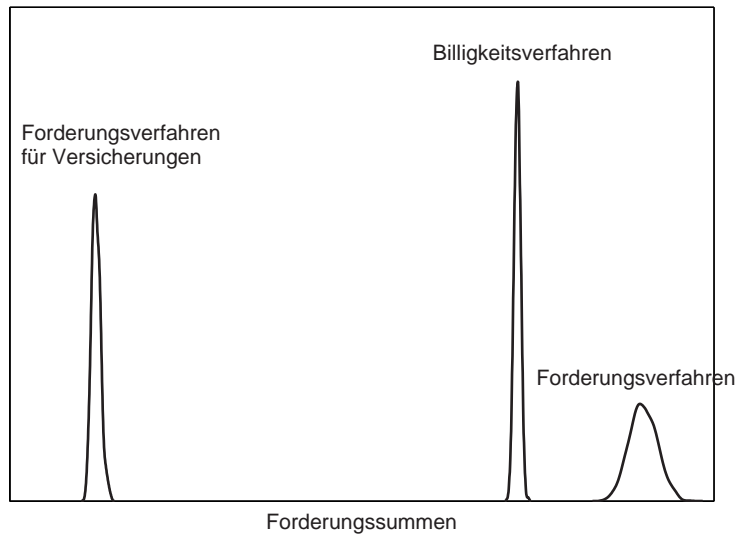


Abbildung 1: Verteilung der Schätzer für die Forderungssummen in den drei Verfahrensarten

5.7 Multiples Imputieren

Nach Ausführung der in Unterkapitel 5.1 bis 5.6 beschriebenen Schritte konnte ein Schätzer für die Forderungssummen berechnet werden. Wir haben diesen Vorgang 2000 mal wiederholt und erhielten damit eine sehr genaue Schätzung für die Verteilung der Schätzer der Forderungssummen.

6 Ergebnisse

6.1 Erster Programmdurchlauf

Für die Modellparameter wurden Einstellungen gewählt, die empirisch gut begründet waren. Die drei Parameter waren:

- Anteil, um den sich die Anzahl der Nacherfassungen bei der Bearbeitung erhöht
- Anteil, um den sich die Anzahl der Ausweitungen bei der Bearbeitung erhöht
- Anteil der tatsächlichen Miterben unter den potentiellen Miterben

In Abbildung 1 sind die Verteilungen für die Schätzer der Forderungssummen zu sehen. Es ist deutlich zu erkennen, dass die Verteilung im Forderungsverfahren am stärksten streut und im Billigkeitsverfahren am wenigsten. Dieses Ergebnis überrascht nicht, waren doch die bearbeiteten und damit bekannten Forderungen im Billigkeitsverfahren unter allen bearbeiteten Forderungen stark überrepräsentiert. Und mehr Daten erlauben einfach eine genauere Schätzung.

Unter der Voraussetzung, dass die obigen Modellannahmen stimmen, kann man sagen, dass die tatsächlichen Forderungssummen in den drei Verfahrensarten mit an Sicherheit grenzender Wahrscheinlichkeit die Maxima der Schätzer für die Forderungssummen nicht überschreiten werden.

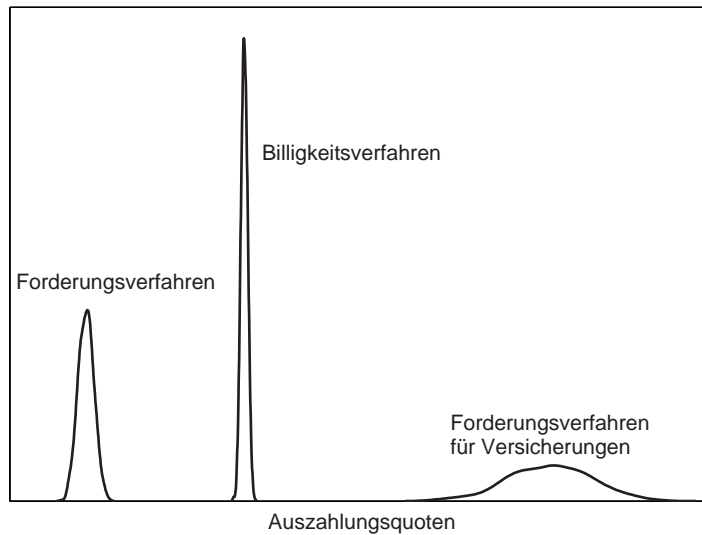


Abbildung 2: Verteilung der Schätzer für die Auszahlungsquoten in den drei Verfahrensarten

Bei der Bestimmung der Schätzer für die Auszahlungsquoten gingen wir bei der ersten Prognose davon aus, dass in den drei Verfahrensarten folgende Beträge zur Ausschüttung kommen:

Billigkeitsverfahren:	92,5 Mio. US \$
Forderungsverfahren:	80 Mio. US \$
Forderungen aus entzogenen Versicherungen:	25 Mio. US \$

(Im Forderungsverfahren hat sich dieser Betrag schließlich auf 92,5 Mio. US \$ geändert.)

Damit ergaben sich für die drei Verfahrensarten die in Abbildung 2 dargestellten Verteilungen für die Schätzer der Auszahlungsquoten. Auch hier kann man wieder deutlich die unterschiedliche Streuung der drei Verteilungen sehen. Die Unsicherheit ist im Forderungsverfahren für Versicherungen am größten, was auch der Grund dafür war, dass schließlich die Vorauszahlungsquote für diese Forderungen gleich hoch wie die Quote für die wesentlich genauer einzuschätzenden Forderungen aus dem Billigkeitsverfahren angesetzt wurden.

Unter der Voraussetzung, dass die obigen Modellannahmen stimmen kann man damit sagen, dass die tatsächlichen Auszahlungsquoten in den drei Verfahrensarten mit an Sicherheit grenzender Wahrscheinlichkeit die Minima der Schätzer für die Auszahlungsquoten nicht unterschreiten werden. Der Median der geschätzten Quoten kann als Punktschätzer für die tatsächlichen Auszahlungsquoten genommen werden.

Auf der Grundlage dieser Quotenschätzung wurden schließlich vom Kuratorium des Nationalfonds der Republik Österreich für Opfer des Nationalsozialismus im Einvernehmen mit dem Antragskomitee des Allgemeinen Entschädigungsfonds folgende Vorauszahlungsquoten an Antragsteller, deren Forderungen bereits entschieden sind,

beschlossen:

- Forderungen, die im Billigkeitsverfahren entschieden wurden: 15% der anerkannten Forderungssumme.
- Versicherungsforderungen, die im Forderungsverfahren entschieden wurden: 15% der anerkannten Forderungssumme.
- alle anderen Forderungen, die im Forderungsverfahren entschieden wurden: 10% der anerkannten Forderungssumme.

6.2 Weitere Programmdurchläufe

Um einerseits die Empfindlichkeit des angewandten Verfahrens auf die Verletzung von Modellannahmen zu testen, und da andererseits die im ersten Durchlauf angenommenen Einstellungen der Programmparameter nicht vollständig theoretisch abgesichert waren, haben wir uns zu zwei weiteren Programmdurchläufen mit geänderten Parametereinstellungen entschlossen.

6.2.1 Höherer Anteil an Miterben

Hier war das Datenmaterial für eine theoretisch fundierte Parameterwahl am schlechtesten. Um die Auswirkungen eines höheren Anteils an Miterben beurteilen zu können, wurde für den zweiten Programmdurchlauf *ceteris paribus* der Anteil der tatsächlichen Miterben unter den potentiellen Miterben verdoppelt.

Wie sich bei den späteren Prognosen mit umfangreicherem Datenmaterial herausstellte, entspricht diese Parametereinstellung viel eher der Realität, und zwar aus folgendem Grund: Der Anteil der bis jetzt tatsächlich ins Verfahren eingebrachten Miterben unter den Antragstellern mit potentiellen Miterben entspricht zwar etwa dem zuerst angenommenen Prozentsatz. Vergleicht man aber die maximal mögliche Summe der Forderungsbeträge der Miterben mit den tatsächlich von den Miterben eingebrachten Forderungssummen, dann ist die Ausschöpfungsquote der Miterben fast doppelt so hoch wie der Anteil der tatsächlich ins Verfahren eingebrachten Miterben an allen möglichen Miterben. Dafür bietet sich folgende Erklärung an: Unter den Forderungen, die für Miterben in Frage kommen, sind viele mit sehr kleinen Forderungsbeträgen. Die Miterben verzichten offensichtlich auf ihren Anspruch, wenn nicht ein gewisser Mindestbetrag übertroffen wird.

Wie sich durch die geänderte Parametereinstellung die Verteilungen der Schätzer für die Auszahlungsquoten verringern, ist in Abbildung 3 zu sehen. Die Verteilungen verschoben sich zwar merklich, doch wirkte sich die Verdopplung der Ausschöpfungsquote der Miterben nicht so dramatisch auf die Minima der Schätzer für die Auszahlungsquoten aus. Die mittleren Schätzer für die Auszahlungsquoten sind auch nur geringfügig niedriger als im ersten Programmdurchlauf.

6.2.2 Höherer Anteil an Nacherfassungen

Hier war das Datenmaterial deutlich aussagekräftiger. Geht man davon aus, dass die Juristen nicht aus irgendwelchen Gründen plötzlich merklich mehr nacherfassen,

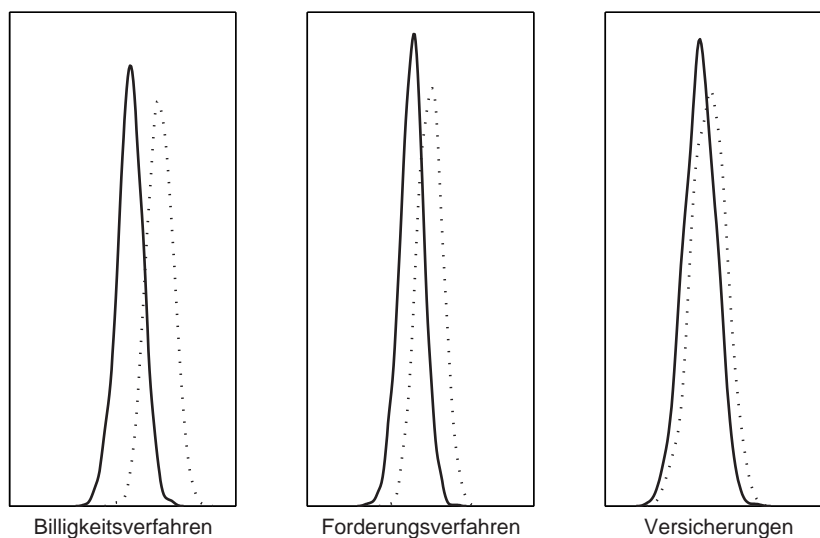


Abbildung 3: Verteilungen der Schätzer für die Auszahlungsquoten mit Ausschöpfung der Miterben bei $2x\%$ bzw. $x\%$ (punktierter Linie)

dann kann man mit großer Sicherheit eine Obergrenze für den Anteil, um den sich die Anzahl der Nacherfassungen bei der Bearbeitung erhöht, angeben. Wir haben für einen dritten Durchlauf diesen Extremwert als Parametereinstellung gewählt, die beiden anderen Programmparameter wurden wie im ersten Durchlauf eingestellt.

Im Nachhinein stellte sich heraus, dass diese Parametereinstellung von geringerer Relevanz ist. Auch mit aktuelleren Daten spricht nichts gegen die Annahme, dass der Anteil, um den sich die Anzahl der Nacherfassungen bei der Bearbeitung erhöht, beim ursprünglich angenommenen Prozentsatz liegt.

Die Auswirkungen dieser Änderung in der Parametereinstellung waren – wie auch zu erwarten war – noch geringer als beim zweiten Durchlauf. Die Forderungssummen erhöhten sich in den drei Verfahrensarten nur um relativ geringfügige Beträge. Auch die Verteilungen der Schätzer für die Auszahlungsquoten verschoben sich, wie in Abbildung 4 zu sehen ist, im Vergleich zum ersten Durchlauf nur unmerklich.

Insgesamt kann man an den zwei weiteren Durchläufen also sehen, dass sich die Verletzung einer der Modellannahmen natürlich auswirkt, die Auswirkungen aber nicht allzu dramatisch sind. Das Verfahren kann aus dieser Sicht als relativ robust gegenüber Verletzungen der Modellannahmen bezeichnet werden.

6.3 Resümee

Mit dem von uns gewählten Verfahren ist es bei einer nicht gerade befriedigenden Datensituation (viele fehlende Daten, wenige Variablen, vorhandene Daten nicht repräsentativ, ...) sehr gut gelungen, genaue Abschätzungen der Verteilungen der Schätzer für die Auszahlungsquoten zu bestimmen. Natürlich ist die Gültigkeit der Ergebnisse von der Gültigkeit der Modellannahmen abhängig. Einem darauf Bezug nehmenden Einwand sind aber zwei Punkte entgegenzuhalten.

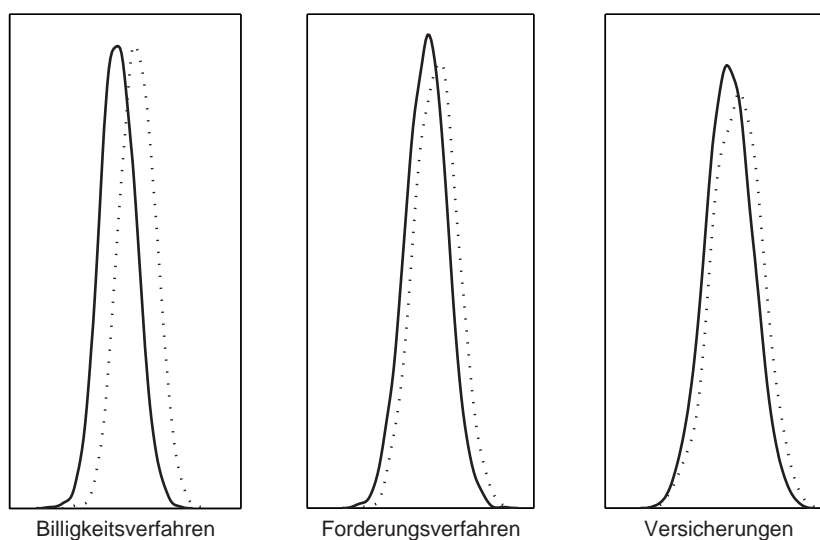


Abbildung 4: Verteilungen der Schätzer für die Auszahlungsquoten mit geringerer (punktierte Linie) bzw. höherer Erhöhung der Nacherfassungen

1. Kein Analyseverfahren kommt ohne Modellannahmen aus, wenngleich bei der Präsentation mancher Untersuchungsergebnisse oft nicht klar ist, welche Annahmen verwendet wurden.
2. Unsere Modellannahmen sind zwar wie alle solche Annahmen Idealisierungen der Realität, stellen aber, soweit es geht, zumindest plausible Schlussfolgerungen aus den vorhandenen Daten dar. Für die eher unsicheren Annahmen wurde außerdem gezeigt, dass das Verfahren nicht labil auf eine Verletzung der Annahmen reagiert.

Dieser ersten Prognose vom November 2005 folgten sechs weitere Quotenprognosen für den Allgemeinen Entschädigungsfonds im April und Oktober 2006 und im Jänner, März, Mai und September 2007. Aufgrund anderer Rahmenbedingungen und einigen Datenänderungen verschoben sich zwar die Verteilungen für die Schätzer für die Auszahlungsquoten geringfügig, die erste, in dieser Arbeit im Detail beschriebene Prognose kann aber im Nachhinein immer noch als brauchbar bezeichnet werden.

Literatur

Allgemeiner Entschädigungsfonds für Opfer des Nationalsozialismus. Homepage: <http://www.de.nationalfonds.org/cgi-bin/dynamic?id=res11>.

Little, R. and D. Rubin (2002). *Statistical Analysis with Missing Data*, 2nd edition. New York: Wiley Series in Probability and Statistics.

Nationalrat der Republik Österreich, BGBl. I Nr. 12/2001. *Bundesgesetz über die Einrichtung eines Allgemeinen Entschädigungsfonds für Opfer des Nationalsozialismus und über Restitutionsmaßnahmen (Entschädigungsfondsgesetz) sowie zur*

Änderung des Allgemeinen Sozialversicherungsgesetzes und des Opferfürsorgegesetzes.

Nationalrat der Republik Österreich, BGBl. I Nr. 142/2005. *Bundesgesetz, mit dem das Entschädigungsfondsgesetz geändert wird.*

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Rubin, D. and N. Schenker (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* 81, 366–374.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.

Autoren

Dipl.-Ing. Mag. Dr. Helmut Waldl, Ass.-Prof. Dr. Andreas Quatember, Univ.-Prof. Dr. Werner G. Müller

alle: Johannes Kepler Universität Linz, Institut für Angewandte Statistik, Altenbergerstraße 69, A-4040 Linz

korrespondierender Autor: Helmut Waldl, e-mail helmut.waldl@jku.at

Werner G. Müller war zur Zeit der Erstellung der Prognose außerordentlicher Universitätsprofessor am Department für Mathematik und Statistik der Wirtschaftsuniversität Wien.