

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
Johannes Kepler Universität Linz

Abstract: The General Settlement Fund for Victims of National Socialism started with the initial payments to applicants at the end of January 2006. In the decisive expertise concerning the advance payments the claims that have not yet been assessed were estimated with an imputation method. These estimates were the basis for fixing the quotas of the initial payments. These quotas amount to 10% or 15% of the recognized claims depending on the categories of property for the various claims for compensation.

Zusammenfassung: Ende Jänner 2006 konnte mit den ersten Auszahlungen des Allgemeinen Entschädigungsfonds (General Settlement Fund, kurz: GSF) begonnen werden. Im entscheidenden Gutachten zu den Auszahlungssummen wurden unter Verwendung einer Imputationsmethode die noch fehlenden Auszahlungssummen geschätzt und auf deren Basis Vorauszahlungsquoten festgelegt. Die Vorauszahlungen betragen je nach Kategorie zehn oder 15 Prozent der als berechtigt anerkannten Forderungen. Die vorliegende Arbeit beschreibt die Vorgehensweise für die erste von mehreren Quotenprognosen, die für den GSF gerechnet wurden.

Keywords: Approximate Bayesian Bootstrap, Multiple Imputation, Vorauszahlungsquoten.

1 Einleitung

Der *Allgemeine Entschädigungsfonds für Opfer des Nationalsozialismus* (General Settlement Fund, kurz: GSF) erbringt seit Jänner 2006 vorläufige Leistungen an Antragsteller. Dazu werden laufend Prognosen der Auszahlungsquoten verbunden mit einer Risikoanalyse als Basis für die Wahl geeigneter Auszahlungsquoten erstellt. Ziel der ersten dieser Analysen 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. Die Beschreibung dieser Analyse ist Gegenstand der vorliegenden Arbeit.

Je nach Vermögenskategorie und Dokumentation der Forderung konnten Entschädigungsansprüche in den unterschiedlichen Verfahrensarten geltend gemacht werden. Einerseits konnten Forderungen nach erleichterten Beweisstandards im sogenannten Forderungsverfahren eingebracht werden. Andererseits konnten Forderungen, die nach den Beweisstandards des Forderungsverfahrens nicht ausreichend dokumentiert waren, im Billigkeitsverfahren geltend gemacht werden, wobei die Bewertung der Verluste pauschaliert

durchgeführt wird. Zusätzlich ist ein eigenes Verfahren für Forderungen aus entzogenen Versicherungen vorgesehen. Für das Billigkeitsverfahren, das Forderungsverfahren und das Verfahren für Forderungen aus entzogenen Versicherungen sollen abhängig von den Vorauszahlungsquoten Wahrscheinlichkeiten für ein Überschreiten der endgültigen Auszahlungsquoten bestimmt werden. Die juristischen Grundlagen für die Vorauszahlungen sind 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.

Bei der Bestimmung der Schätzer für die Auszahlungsquoten gehen 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 änderte sich der Betrag schließlich auch auf 92.5 Mio. US\$. Diese Beträge sind dem Entschädigungsfondsgesetz (BGBl. I Nr.12/2001, §2 bzw. §5) zu entnehmen.

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.

Basis der Analyse ist eine vom Fonds zur Verfügung gestellte, laufend aktualisierte MS-Access Datenbank. In Kapitel 2 dieses Aufsatzes findet man die theoretischen Grundlagen der verwendeten statistischen Methoden. Die konkrete Vorgehensweise wird im 3. Kapitel erläutert. Kapitel 4 beschreibt die Ergebnisse, soweit sie veröffentlicht werden dürfen.

2 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 Faktum. Spezielle Methoden wie die Imputationsmethoden wurden entwickelt, die auf diese Problematik eingehen (für einen Überblick zur Problematik siehe etwa Särndal, Swensson, und 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 set“) an Fällen, für die das interessierende Merkmal y , die Forderungshöhe, bereits bestimmt wurde, und eine Teilmenge M („missing set“) 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, so lässt sich \hat{t} durch Summation über alle Schichten darstellen:

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

($h = 1, \dots, K$) mit der Menge U_h der Fälle in der h -ten Schicht. Diese lässt sich wiederum in eine Responsemenge R_h und eine Missingmenge M_h zerlegen.

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. Eine Single Imputations-Methode ermöglicht jedoch keine weder eine korrekte direkte Schätzung der Verteilung noch eine solche der Streuung des Schätzers \hat{t} . Im vorliegenden Sonderfall, in dem die fehlenden Werte in U selbst und nicht in einer Stichprobe aus U auftreten, würde die geschätzte Varianz fälschlicherweise sogar Null betragen.

Die *multiple Imputation* von gleich D Ersatzwerten pro fehlendem Wert hingegen ermöglicht eine Verteilungsschätzung, die den Imputationsfehler als Streuungsquelle mitberücksichtigt, indem durch die Erzeugung von D Datensätzen zusätzlich zum Stichprobenfehler auch die rein durch die Imputationen verursachte Streuung zwischen den daraus resultierenden D Schätzern \hat{t} berücksichtigt werden kann (siehe dazu Rubin, 1987). Da bei multipler Imputation von Ersatzwerten mittels 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. Dieser Schritt spiegelt die Unsicherheit der theoretischen Verteilung von y , aus der gezogen werden soll, wider. Das Verfahren wird als *Approximate Bayesian Bootstrap* bezeichnet (siehe Little und 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.

3 Beschreibung der Analyse

Vor der eigentlichen Analyse werden bei der Datenaufbereitung statistische Zusammenhänge zwischen Variablen entdeckt, die bei der späteren Schichtung der Forderungen

berücksichtigt werden. Für fehlende Daten (die Daten der nicht bearbeiteten Antragsteller und der noch nicht bearbeiteten Forderungen) werden vorhandene Daten von Antragstellern bzw. Forderungen mit ähnlicher Struktur eingesetzt. Beispielsweise wird bei der Forderung eines österreichischen Antragstellers mit Forderungskategorie Betriebe, die sich auf eine weibliche historische Person bezieht, bei der die Forderungssumme fehlt, zufällig eine Forderungssumme aus jenen vollständig vorhandenen Datensätzen gezogen, bei denen der Antragsteller „Österreicher“, die Kategorie „Betriebe“ und die historische Person „weiblich“ ist.

Da die wiederholte Datenimputation sehr rechenintensiv ist, kann man sie nicht beliebig oft durchführen. Da wir für die Prognose die gesamte Verteilung der Schätzer und insbesondere die niedrigen Quantile benötigen, und nicht wie häufig üblich nur einzelne Parameter der Verteilung schätzen wollen, entscheiden wir uns für eine hohe Wiederholungszahl von 2000.

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. Zur Unterstreichung ihrer Wichtigkeit kennzeichnen wir die getroffenen Annahmen als solche. Im Wesentlichen handelt es sich dabei um die Annahme „missing at random“ hinsichtlich des Nonresponse-Mechanismus (vergleiche Rubin, 1987), welche für die folgenden Variablen gilt:

- Geschlecht der historischen Person
- Erbquote
- Summe der Erbquoten der zugehörigen historischen Person
- Vermögenskategorie des Verlustes
- vom Antragsteller gewünschte Verfahrensart
- Verfahrensart, in der die Forderung entschieden wurde
- Ausweitung
- nacherfasste Forderung

Dies wird in der Folge noch im Detail beschrieben.

3.1 Imputation nicht erfasster Antragsteller

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 erfolgt in mehreren Schritten. Zuerst muss für 6207 nicht erfasste Antragsteller eine Forderungsstruktur (die Menge aller Forderungen eines Antragstellers) imputiert werden. Die Forderungsstruktur wird bei der ersten Quotenprognose aus 12967 Antragstellern gezogen, die entweder erfasst waren oder bereits eine Forderungsstruktur besitzen. Von diesen 12967 Antragstellern sind 7127 nicht bearbeitet.

Die ohnehin nur in geringer Anzahl vorhandenen Antragsteller, die erfasst sind, aber keine zugehörige historische Person und keine Forderung hatten, wird entfernt. Ebenso wird bei wenigen Antragstellern jeweils eine historische Person gelöscht, zu der keine Forderungen existieren. Bei der Imputation der Forderungsstrukturen berücksichtigen wir die Tatsache, dass unter den bearbeiteten Antragstellern jene mit geringer Forderungszahl und jene mit Forderungen der Kategorie „berufs- und ausbildungsbezogene Verluste“ stark überrepräsentiert sind.

Modellannahme 1: Es wird 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 sind die bearbeiteten Antragsteller daher entsprechend zu gewichten. Für die imputierten Antragsteller übernehmen wir bei allen Forderungen die Werte der folgenden Variablen:

- Vermögenskategorie des Verlustes:
 - Aktien, Wertpapiere
 - Bankguthaben
 - Beruf, Ausbildung
 - Betrieb
 - bewegliches Vermögen
 - Hypothekarforderung
 - Immobilien
 - Schuldverschreibungen
 - sonstige Verluste
 - Versicherungen
- vom Antragsteller gewünschte Verfahrensart:
 - Billigkeitsverfahren
 - Forderungsverfahren
 - keine Verfahrensart gewählt
- Geschlecht der historischen Person:
 - männlich
 - neutral (Vereine, Organisationen)
 - weiblich
- Ausweitung:
 - ja
 - nein
- nacherfasste Forderung:
 - ja
 - nein

Die beiden Variablen „Ausweitung“ und „nacherfasste Forderung“ kennzeichnen Forderungen, die den Anträgen nachträglich hinzugefügt werden. Bei Ausweitungen handelt es sich um Verluste, die durch die Recherche des GSF gefunden und vom Antragsteller nicht selbst geltend gemacht werden. Der Großteil dieser Ausweitungen findet im Stadium der Datenerfassung statt. Nacherfasste Forderungen erfasst man erst, wenn die Abteilung

Datenerfassung des GSF den Akt als fertig erfasst gekennzeichnet hat. Diese Forderungen entstehen meist dadurch, dass die Recherche noch zusätzliche Dokumente zu Tage fördert, nachdem die Datenerfassung abgeschlossen ist.

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

Nicht übernommen wird der Wohnsitzstaat des Antragstellers (er ist auch bei nicht erfassten Antragstellern bekannt) und die Bewertung der Forderung. Außerdem wird unterschieden, ob aus bearbeiteten oder aus nur erfassten Antragstellern imputiert wird. Bei den Imputationen aus den bearbeiteten Antragstellern werden auch die folgenden Variablen übernommen:

- Verfahrensart, in der die Forderung entschieden wurde:
 - Billigkeitsverfahren
 - Forderungsverfahren
- Erbquote
- Summe der Erbquoten der zugehörigen historischen Person

Außerdem werden 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 noch Forderungen durch Ausweitung und Nacherfassung hinzukommen bzw. durch Verwerfen verschwinden.

3.2 Simulation von Nacherfassungen und Ausweitungen

3.2.1 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 innerhalb einiger Kategorien ein sonderbarer Effekt zu beobachten: Besonders bei den Kategorien „Immobilie“ und „Schuldverschreibungen“ ist 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 gilt somit: 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.

3.2.2 Ausweitungen

Hier ist die Situation ähnlich der bei den nacherfassten Forderungen. Die Anteile der Ausweitungen sind unter den Forderungen der bearbeiteten Antragsteller und unter jenen der nur erfassten Antragsteller etwa gleich. Besonders in den Kategorien „Immobilien“, „sonstige Verluste“ und „Versicherungen“ ist 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 vorkommen. 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 wird 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 wird, stellt die ungleiche Verteilung der Ausweitungen und Nacherfassungen unter den nur erfassten bzw. unter den fertig bearbeiteten Antragstellern kein Problem dar. Hinsichtlich der zusätzlichen Ausweitungen und Nacherfassungen erfolgen während der juristischen Bearbeitung nach Auskunft des Projektkoordinators des Allgemeinen Entschädigungsfonds für Opfer des Nationalsozialismus, Mag. Hubinger, „... Ausweitungen bei der juristischen Bearbeitung nur sporadisch ...“ und es wird nur „... ein kleiner Teil der nacherfassten Forderungen von den Juristen erzeugt ...“. Wie groß dieser Anteil ist, konnte nicht exakt quantifiziert werden.

Modellannahme 4: Es wird davon ausgegangen, dass sich während der Bearbeitung der Akte die Anzahl der Ausweitungen und die Anzahl der Nacherfassungen um Prozentsätze erhöhen, die aufgrund der fertig bearbeiteten Akten plausibel erscheinen und im Vorhinein bestimmt werden. Da diese Annahmen eher vage sind, sind die obigen Prozentsätze Programmparameter, die für neue Programmdurchläufe jederzeit geändert werden können.

Im Programm wird 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 werden.

3.3 Simulation der Verwerfungen

Bei den verworfenen Forderungen sind starke Abhängigkeiten zu beobachten: So werden etwa von den nacherfassten Forderungen weniger verworfen als vom Rest, oder von den Ausweitungen mehr als vom Rest. Der Anteil der verworfenen Forderungen ist 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 wird 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 wird als zuvor) und dass die Forderungen der bearbeiteten Antragsteller die Anteile der verworfenen Forderungen nach der juristischen Bearbeitung für die „zukünftige“ Aktenerledigung gut repräsentieren.

Bei den nicht entschiedenen und den imputierten Forderungen wird entsprechend den Anteilen der verworfenen Forderungen in den fertig bearbeiteten Akten zufällig verworfen, und zwar geschichtet nach folgenden Variablen:

- Vermögenskategorie des Verlustes
- Ausweitung
- nacherfasste Forderung

3.4 Simulation der Änderung der Verfahrensart

Nach dem Verwerfen von Forderungen muss für alle nicht entschiedenen Forderungen mit Ausnahme jener, die aus bearbeiteten Antragstellern imputiert werden, die Verfahrensart, in der über die Forderung entschieden wird, simuliert werden. Bei der Datenanalyse ist unter den entschiedenen Forderungen eine klare Abhängigkeit der Verfahrensart, in der die Forderung entschieden wird, von der Vermögenskategorie des Verlustes, von der Variable Ausweitung und der vom Antragsteller gewünschten Verfahrensart festzustellen. So wechselt zum Beispiel die Verfahrensart bei Ausweitungen kaum, da die vom Antragsteller gewünschte Verfahrensart vom Bearbeiter des Aktes bestimmt wird.

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

Die Verfahrensart, in der die Forderung entschieden wird, wird entsprechend den unter den entschiedenen Forderungen gefundenen Anteilswerten dieser Variablen simuliert. Dabei wird nach folgenden Variablen geschichtet:

- Vermögenskategorie des Verlustes
- Ausweitung
- vom Antragsteller gewünschte Verfahrensart

Bei den Forderungen, die aus bearbeiteten Antragstellern imputiert werden, ist das nicht notwendig. Für sie ist die Verfahrensart bekannt, in der die Forderung entschieden wird.

3.5 Imputation der Bewertung der Forderungen

Als Nächstes wird für die nicht bearbeiteten und imputierten Forderungen die Bewertung der Forderungen aus abgelehnten und zuerkannten Forderungen imputiert. Da 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, wird die Menge der abgelehnten und zuerkannten Forderungen nach diesen Variablen geschichtet, und zwar so, dass in

jeder Schicht eine Mindestzahl von 100 Beobachtungen vorhanden ist. Bei den meisten Schichten ist die Anzahl an Beobachtungen aber wesentlich größer. Insgesamt wurde aus 80 verschiedenen Schichten nach dem in Abschnitt 2 beschriebenen Verfahren imputiert. Die verschiedenen Ausprägungen der Variablen, die zur Schichtung verwendet werden, sind der folgenden Aufstellung zu entnehmen bzw. bereits weiter oben angeführt:

- Vermögenskategorie des Verlustes
- Verfahrensart, in der die Forderung entschieden wurde
- Wohnsitzstaat des Antragstellers:
 - Australien, Neuseeland
 - Israel
 - Österreich
 - USA, Kanada
 - restliche Staaten
- Geschlecht der historischen Person

Modellannahme 7: Es wird 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.

3.6 Simulation von Miterben

Miterben können bei historischen Personen vorkommen, wenn die Summe der Erbquoten aller Antragsteller mit dieser historischen Person < 1 ist. Unter den entschiedenen Antragstellern sei dies bei $x\%$ der Fall. Dieser Anteil verringert sich bei Berücksichtigung noch nicht bearbeiteter Antragsteller.

Modellannahme 8: Es wird 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 wird 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 ist, wird der Anteil der tatsächlichen Miterben unter den Antragstellern, welche Miterben einbringen können, für einen weiteren Programmdurchlauf ebenfalls zu einem frei wählbaren Parameter. Die Miterben werden 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 werden.

4 Ergebnisse

4.1 Analyse

Für die Modellparameter werden Einstellungen gewählt, die empirisch gut begründet sind. Die drei Parameter sind:

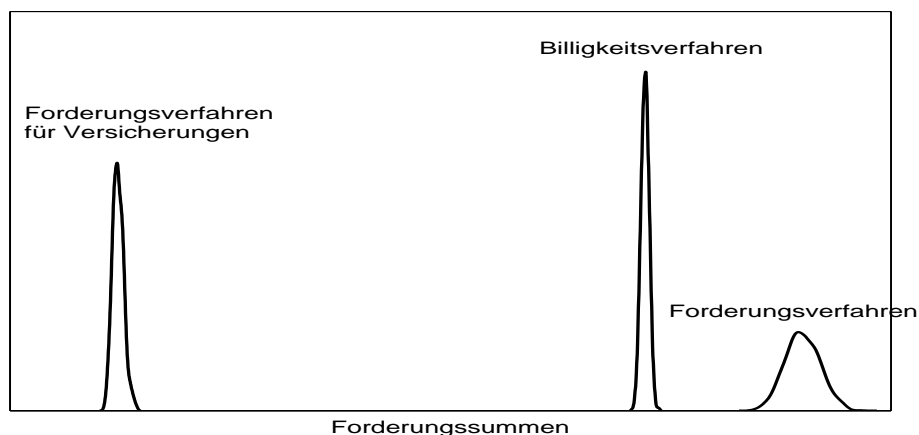


Abbildung 1: Verteilung der Schätzer für die Forderungssummen in den Verfahrensarten

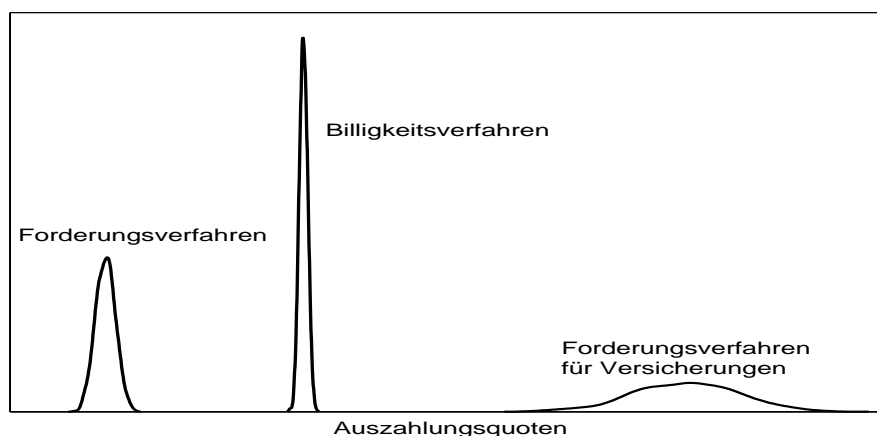


Abbildung 2: Verteilung der Schätzer für die Auszahlungsquoten in den Verfahrensarten

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

Die Setzung der Parameter sowie das Weglassen der Skalierung in den Grafiken erfolgte wegen der uns auferlegten Geheimhaltungspflicht. 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. Dies ist darauf zurückzuführen, dass die bearbeiteten und damit bekannten Forderungen im Billigkeitsverfahren unter allen bearbeiteten Forderungen stark überrepräsentiert waren. Mehr Daten erlauben natürlich 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. Damit ergeben sich für die drei Verfahrensarten die in Abbildung 2 dargestellten Verteilungen für die Schätzer der Auszahlungsquoten. Auch hier kann man 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 (siehe Kapitel 1).

Unter der Voraussetzung, dass die obigen Modellannahmen stimmen, kann gesagt werden, dass die tatsächlichen Auszahlungsquoten in den drei Verfahrensarten mit an Sicherheit grenzender Wahrscheinlichkeit die Minima der Schätzer für die Auszahlungsquoten nicht unterschreiten. Der Median der geschätzten Quoten kann als Punktschätzer für die tatsächlichen Auszahlungsquoten herangezogen 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 (siehe Allgemeiner Entschädigungsfonds für Opfer des Nationalsozialismus, 2007):

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

4.2 Sensitivitätsanalyse

Um die Empfindlichkeit des angewandten Verfahrens auf die Verletzung von Modellannahmen zu testen und weil die im ersten Durchlauf angenommenen Einstellungen der Programmparameter nicht vollständig theoretisch abgesichert sind, wurden zwei weitere Programmdurchläufe mit geänderten Parametereinstellungen durchgeführt.

4.2.1 Höherer Anteil an Miterben

Hier ist das Datenmaterial für eine theoretisch fundierte Parameterwahl am schlechtesten. Um die Auswirkungen eines höheren Anteils an Miterben beurteilen zu können, wird 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 eher der Realität. Die Begründung dafür ist, dass der Anteil der bis dato tatsächlich ins Verfahren eingebrachten Miterben unter den Antragstellern mit potentiellen Miterben wohl etwa dem zuerst angenommenen Prozentsatz entspricht. 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.

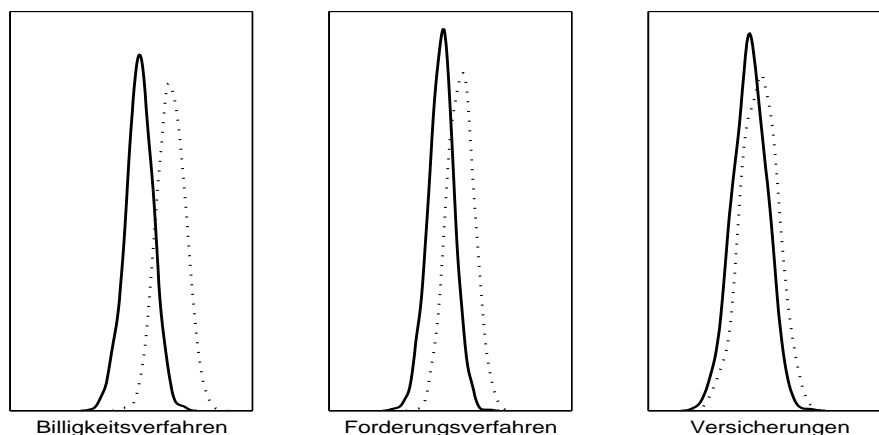


Abbildung 3: Verteilungen der Schätzer für die Auszahlungsquoten mit Ausschöpfung der Miterben bei $2x\%$ bzw. $x\%$ (punktierte Linie)

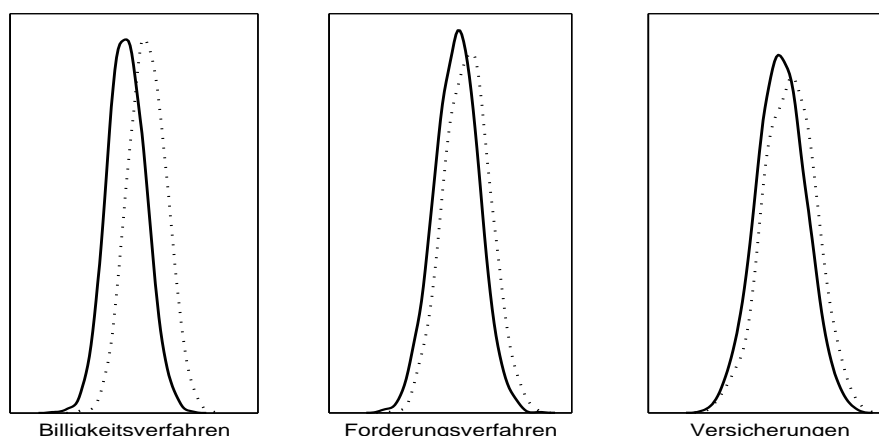


Abbildung 4: Verteilungen der Schätzer für die Auszahlungsquoten mit geringerer (punktierte Linie) bzw. höherer Erhöhung der Nacherfassungen

Die Veränderung der Verteilungen der Schätzer für die Auszahlungsquoten ist in Abbildung 3 dargestellt. Die Verteilungen verschieben sich zwar, doch wirkt sich die Verdopplung der Ausschöpfungsquote der Miterben nicht 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.

4.2.2 Höherer Anteil an Nacherfassungen

Hier ist das Datenmaterial deutlich aussagekräftiger. Geht man davon aus, dass nicht plötzlich merklich mehr nacherfasst werden wird, dann lässt sich 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 genau so wie im ersten Durchlauf eingestellt.

Die Auswirkungen dieser Änderung in der Parametereinstellung sind noch geringer als beim zweiten Durchlauf. Die Forderungssummen erhöhen sich in den drei Verfahrensarten nur um relativ geringfügige Beträge. Auch die Verteilungen der Schätzer für die Auszahlungsquoten verschieben sich, wie in Abbildung 4 dargestellt wird, im Vergleich zum ersten Durchlauf nur unmerklich.

Weitere Sensitivitätsanalysen wurden mit anderen Parametereinstellungen durchgeführt, die Ergebnisse ähnelten den oben dargestellten Vergleichen. Auf darüber hinausgehende Analysen (z.B. bezüglich der Repräsentativitätsannahmen) musste aufgrund der umfangreichen Datenaufbereitung und der langen Rechenzeit verzichtet werden. Man kann an den zwei weiteren Durchläufen erkennen, dass sich die Verletzung einer der Modellannahmen nicht dramatisch auswirkt. Das Verfahren darf demnach als relativ robust gegenüber Verletzungen der Modellannahmen bezeichnet werden.

5 Resümee

Mit dem von uns gewählten Verfahren ist es bei einer unbefriedigenden Datensituation (viele fehlende Daten, wenige Variablen, vorhandene Daten nicht repräsentativ und so fort) offenbar gut gelungen, genaue Abschätzungen der Verteilungen der Schätzer für die Auszahlungsquoten zu bestimmen. Dies lässt sich an der mittlerweile stark vergrößerten Datenbasis nachweisen (sieben weitere Quotenprognosen im April und Oktober 2006 und im Jänner, März, Mai, September und November 2007), wenngleich vor dem Abschluss der Arbeiten des GSF kein endgültiges Urteil möglich ist. Natürlich ist die Gültigkeit der Ergebnisse von der Gültigkeit der Modellannahmen abhängig. Einem darauf Bezug nehmenden eventuellen Einwand sind aber zwei Punkte entgegenzuhalten:

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.

Für die Erstellung der Quotenprognose wurde folgende Software verwendet:

- Microsoft Office Access 2003
- Microsoft Office Excel 2003
- SPSS 14.0 für Windows
- MATLAB Version 7.1.0.246

Die Hauptarbeit wurde mit MATLAB, einem Hochleistungsprogrammpaket für numerische Berechnungen durchgeführt, die anderen Programme wurden zur Datenaufbereitung verwendet. Die Rechenzeit für die 2000malige Simulation des gesamten Fonds betrug auf einem Standardarbeitsplatz-Computer der Johannes Kepler Universität etwas mehr als 10 Stunden.

Anerkennung

Die Autoren bedanken sich bei den Gutachtern für die ausführlichen Kommentare und Anregungen.

Literatur

- Allgemeiner Entschädigungsfonds für Opfer des Nationalsozialismus. (2007). *Vorauszahlungen*. (<http://www.de.nationalfonds.org/cgi-bin/dynamic?id=res11> besucht am 30. November 2007.)
- Little, R. J. A., und Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: Wiley Series in Probability and Statistics.
- Nationalrat der Republik Österreich, BGBl. I Nr. 12/2001. (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. (2005). *Bundesgesetz, mit dem das Entschädigungsfondsgesetz geändert wird*.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.
- Rubin, D. B., und Schenker, N. (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., Swensson, B., und Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.

Adresse der Autoren:

Helmut Waldl,
Andreas Quatember und
Werner G. Müller
IFAS – Institut für Angewandte Statistik
Johannes Kepler Universität Linz
Altenbergerstraße 69
A-4040 Linz
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.