

# Verteilungsformen der Einlagenbestände auf Sparkonten

Von Bernhard Korte und Walter Oberhofer, Bonn

## 1. Vorbemerkungen

Die vorliegende Arbeit stellt einen Beitrag zur quantitativen Analyse des mikroökonomischen Sparprozesses dar. Hier soll versucht werden, mit ökonometrischen und statistischen Methoden zusätzliche und bisher zum Teil unbekannte Aussagen über die Entwicklung, Bestimmungsgründe und die soziale Struktur des Sparprozesses im Hinblick auf den einzelnen Sparer zu gewinnen. Die nachfolgenden Aussagen erheben daher nicht den Anspruch, primär neue Erkenntnisse zur Theorie des Sparens zu liefern. Andererseits war es auch nicht unsere Aufgabe, ein vollständiges und a priori gegebenes theoretisches Sparmodell zu verifizieren. Der Ausgangspunkt unserer Untersuchungen bestand vielmehr darin, daß wir relativ umfangreiches und in dieser Form bisher noch nicht vorhandenes empirisches Material zur Verfügung hatten, aus dem dann entsprechende Resultate deduziert wurden. Diese Erkenntnisse können dann unter Umständen als weitere Bausteine in ein entsprechendes theoretisches Gebäude eingesetzt werden.

Wir wollen ferner bereits hier erwähnen, daß sich unsere Untersuchungen ausschließlich auf das Kontensparen beziehen, da für die anderen Sparformen das notwendige empirische Material unseres Wissens bisher noch nicht vorhanden ist. Diese Spezialisierung bedeutet aber insofern keine allzu starke Einschränkung, als zum Beispiel nach Berechnungen von R. Exo rund 50 % der privaten Ersparnisbildung sowohl der Unselbständigen als auch der Selbständigen auf Sparkonten erfolgt<sup>1</sup>. Diese Prozentzahl hat auch dann Gültigkeit, wenn man als Bezugsgröße den privaten Haushalt nimmt. Sie wurde von R. Exo nicht nur für eine kurze Zeitperiode, sondern insgesamt für 11 Jahre von 1950 bis 1960

---

<sup>1</sup> Vgl. R. Exo, Die Entwicklung der sozialen und ökonomischen Struktur der Ersparnisbildung in der Bundesrepublik Deutschland, Berlin, 1967. Tabellen 53, 54 u. 55, Seite 326 ff.

berechnet. Der obengenannte Wert von 50 % stellt einen Mittelwert über die genannten Jahre dar. Insofern beziehen sich unsere Untersuchungen auf die dominierende Komponente des privaten Sparprozesses.

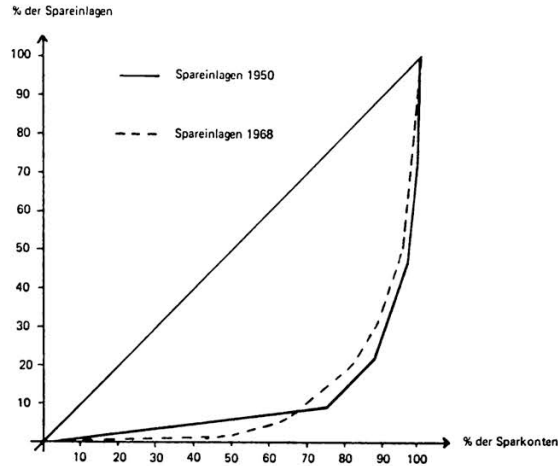
Die bisherigen Analysen des Sparprozesses untersuchten im wesentlichen die *Durchschnittswerte* der Einlagenbestände oder der Umsätze im Verlauf der Zeit bzw. für verschiedene soziale Gruppen. Aus der Veränderung und den Unterschieden der Durchschnittswerte wurden dann entsprechende Schlüsse gezogen. Allenfalls wurden noch grobe Größenklassengliederungen für die Einlagenbestände zur Analyse hinzugezogen. Derartige Durchschnittsbetrachtungen sagen aber nichts über den Typ und die spezielle Form der Verteilungsfunktion der einzelnen Variablen aus. Die Fragen nach der *individuellen Verteilung* der Einlagenbestände und ihrer Parameter (Mittelwert, Varianz) auf Sparkonten konnten daher bisher nicht beantwortet werden. Für eine differenzierte Analyse des Sparprozesses interessieren aber nicht nur die Veränderungen des Mittelwertes, sondern vielmehr Aussagen über mögliche Veränderungen der gesamten Verteilungsfunktion oder der Verteilungsdichte. Ebenso können vergleichende Betrachtungen zwischen verschiedenen sozialen Gruppen, die sich lediglich auf Mittelwerte beschränken, keine signifikanten Unterschiede zeigen, während entsprechende Analysen der Verteilungsfunktion und der Verteilungsdichte zusätzliche Erkenntnisse vermitteln können. Aber neben diesen theoretischen Aussagen hat eine derartige Analyse der Verteilungsformen auch eine nicht geringe praktische Bedeutung. Insbesondere bei der Festlegung von Größenklassen für laufende Statistiken des Sparprozesses ist eine genaue Kenntnis der Verteilungsfunktion wesentlich. Hierdurch kann man vermeiden, daß durch falsche (z. B. zu grobe oder zu feine) Größenklassenbildungen entsprechende Teile der gesamten Verteilung aggregiert, verzerrt oder falsch wiedergegeben werden.

Es sei in diesem Zusammenhang noch eine heuristische Methode zur Beurteilung von Verteilungsformen genannt, bei der eine genaue Kenntnis der Verteilungsfunktion bzw. der Dichtefunktion nicht notwendig ist. Hierbei handelt es sich um die von *Lorenz* zum ersten Mal benutzte Kurve zur Messung des Konzentrationsgrads<sup>2</sup>. In Abbildung 1 sind zwei derartige *Lorenz-Kurven* für die Verteilung der Spareinlagen auf Sparkonten wiedergegeben. Der Graph der *Lorenz-Kurve* entsteht dadurch, daß man jeweils den kumulierten relativen Anteil der Sparkonten an

<sup>2</sup> *Lorenz*, Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association 9 (New Series), (1905), S. 205—219.

Abbildung 1

**Lorenz-Kurve: Konzentrationsmaß für die Verteilung der Spareinlagen auf Sparkonten**



der Gesamtheit aller Sparkonten dem kumulierten relativen Anteil der Spareinlagen auf diesen Sparkonten ebenfalls gemessen an der Gesamtheit aller Spareinlagen gegenübergestellt. Hierzu benutzt man meistens eine Spareinlagen-Größenklassengliederung für die Sparkonten und die Spareinlagen. Nur dann, wenn auf allen Sparkonten jeweils derselbe Spareinlagenbestand vorhanden wäre, würde die entsprechende *Lorenz-Kurve* der in Abbildung 1 eingezeichneten Winkelhalbierenden entsprechen. Denn in diesem Fall würden z. B. auf 10 % aller Sparkonten auch 10 % der Spareinlagen, auf 20 % der Sparkonten auch 20 % der Spareinlagen usw. vorhanden sein. Eine derartige identische Verteilung wird aber in der Regel bei keiner wirtschaftlichen Variablen anzutreffen sein. Daher ist der Graph der *Lorenz-Kurve*, die insbesondere auch zur Analyse der personellen Einkommensverteilung benutzt wird, konkav und unterhalb der Winkelhalbierenden verlaufend. So zeigt z. B. die Abbildung 1, daß im Jahr 1950 auf 50 % aller Sparkonten nur rund 6 % der Spareinlagen vereinigt waren, während z. B. 50 % der Spareinlagen auf nur insgesamt 4 % der Sparerkonten verteilt waren. Die entsprechende *Lorenz-Kurve* für das Jahr 1968 zeigt mit Ausnahme des unteren Bereichs einen gewissen *Dekonzentrationseffekt*. Dieser Effekt ist optisch dadurch festzustellen, daß die entsprechende *Lorenz-Kurve* im



überwiegenden Teil oberhalb der Kurve für das Jahr 1950 verläuft. Das bedeutet also, daß im Verlauf der Zeit die Verteilung der individuellen Spareinlagen auf den Sparkonten etwas gleichmäßiger geworden ist. Diese Aussage stellt aber das Maximum an Information dar, das man mit dem heuristischen Instrument einer *Lorenz*-Kurve ermitteln kann. Eine vertiefte Analyse ist nur bei Kenntnis der Verteilungsfunktion möglich.

Die nächste Frage, der wir uns nun zuwenden wollen, lautet: Welche Verteilungsfunktion kann man *theoretisch* für den Einlagenbestand auf einem einzelnen Sparkonto ableiten? Wir wollen im folgenden mit  $B_i$  den Einlagenbestand auf einem einzelnen Sparkonto am Ende des  $i$ -ten Jahres bezeichnen.

Unterstellt man nun für den Einlagenbestand  $B_i$ , daß er vom Wert der Vorperiode  $B_{i-1}$  abhängt, wobei eine weitere Zufallsvariable als Störterm hinzugenommen wird, so kann man folgenden Ansatz formulieren:

$$(1.1) \quad B_i = B_{i-1} (1 + z_i) .$$

Diese Formel besagt, daß der Einlagenbestand am Ende einer Periode proportional zum Einlagenbestand der Vorperiode ist, wobei der Proportionalitätsfaktor durch die zufällige Größe  $z_i$  angegeben wird. Die relative stochastische Abhängigkeit des Einlagenbestands von demjenigen der Vorperiode erscheint sehr sinnvoll. Aus Formel (1.1) folgt dann für den Einlagenbestand der  $n$ -ten Periode:

$$(1.2) \quad B_n = B_0 (1 + z_1) (1 + z_2) \dots (1 + z_n) ,$$

wobei  $B_0$  der Einlagenbestand zu Beginn eines Beobachtungszeitraums ist. Durch Logarithmieren von (1.2) erhält man:

$$(1.3) \quad \log B_n = \log B_0 + \log (1 + z_1) + \log (1 + z_2) + \dots + \log (1 + z_n) .$$

Nun folgt aus dem *zentralen Grenzwertsatz* der Wahrscheinlichkeitsrechnung, daß die auf der linken Seite von Formel (1.3) stehende Größe  $\log B_n$  *asymptotisch normal* verteilt ist, wobei für die zufälligen Größen  $z_i$  nur sehr schwache Voraussetzungen gefordert werden. So gilt die Aussage des zentralen Grenzwertsatzes, wenn alle  $z_i$  statistisch unabhängig sind und einer beliebigen aber gleichen Verteilungsfunktion mit endlichem Mittelwert und endlicher Streuung genügen. Es lassen sich aber noch wesentlich allgemeinere Voraussetzungen für die Gültigkeit des zen-

tralen Grenzwertsatzes und damit die Gültigkeit der Aussage, daß die Größe  $\log B_n$  normalverteilt ist, angeben. Mit Hilfe der obigen, im allgemeinen sehr sinnvollen Annahme kann man daher theoretisch ableiten, daß der Spareinlagenbestand *logarithmisch normalverteilt* ist, d. h. die Verteilungsdichte der Größe  $B_n$  hat die Form

$$(1.4) \quad f(x) = \frac{1}{x \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (\log x - \mu)^2 \right\}.$$

Hierbei sind  $\mu$  und  $\sigma$  die Parameter (Mittelwert und Streuung) der entsprechenden *Normalverteilung* von  $\log B_n$ .

Mit einem ähnlichen Ansatz hat R. Gibrat zum ersten Mal abgeleitet, daß das individuelle Einkommen ebenfalls logarithmisch normalverteilt ist<sup>3</sup>. Dieses theoretische Resultat ist für alle bekannten Einkommensverteilungen statistisch sehr gut zu verifizieren. Unsere nachfolgenden Ausführungen werden zeigen, daß eine entsprechende Verifikation für die Verteilung der Spareinlagenbestände ebenfalls möglich ist.

Als statistisches Ausgangsmaterial für unsere Untersuchungen haben wir die Daten einer *Stichprobenerhebung* benutzt, die mit einem Umfang von 12 413 Sparkonten im Auftrag des Deutschen Sparkassen- und Giroverbands im Jahre 1964 bei allen öffentlichen Sparkassen durchgeführt wurde. Für jedes dieser Sparkonten wurden die Bestände für einen Zeitraum von 5 Jahren sowie zusätzliche Merkmale über die Person des Kontoinhabers ermittelt. Eine genaue Beschreibung der Art und des Umfangs der Stichprobe sowie der einzelnen Erhebungsmerkmale findet der Leser in einer früher publizierten Arbeit<sup>4</sup>.

## 2. Verteilungsfunktion

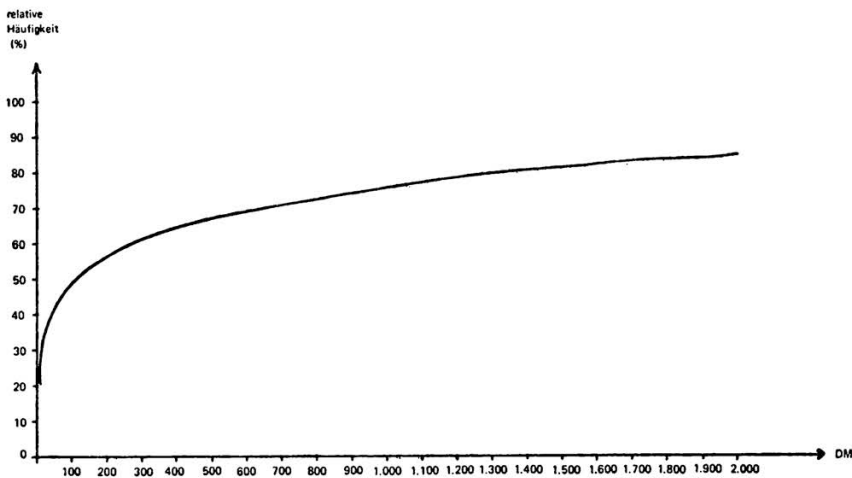
Wir wollen nun aus den empirischen Daten Aussagen über die Verteilungsfunktion der Einlagenbestände auf Sparkonten ableiten. Zu diesem Zweck betrachten wir die Einlagenbestände auf den einzelnen Sparkonten jeweils am Jahresende in der Zeit von 1959 bis 1963, die wir mit  $B_{1959+i}$  ( $i = 0, 1, \dots, 4$ ) bezeichnen wollen. Diese Größen werden im folgenden als sogenannte *zufällige Größen* angesehen, deren Verteilungsfunktion bzw. Verteilungsdichte zu bestimmen ist. Insofern ist also zum Beispiel der Einlagenbestand am Jahresende 1959 auf einem ein-

<sup>3</sup> R. Gibrat, *Les inégalités économiques*. Paris 1931.

<sup>4</sup> Vgl. B. Korte, Stichprobenverfahren und Auswertungstechnik, in: Sparkasse, 83. Jahrgang, Heft 2/3, Februar 1966, S. 46 ff.

zeln Sparkonto eine Realisierung der zufälligen Größe  $B_{1959}$ . Auf Grund der obengenannten Stichprobe haben wir dann 12 413 Realisierungen dieser zufälligen Größen.

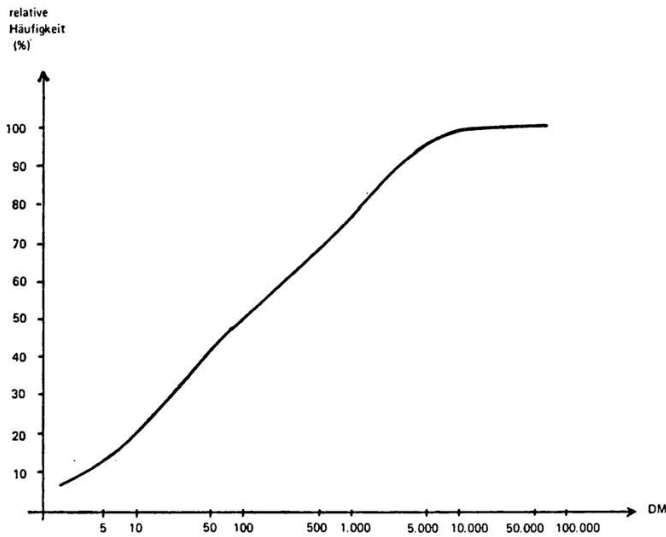
*Abbildung 2*  
**Empirische Verteilungsfunktion des Spar-**  
**einlagenbestands am Jahresende 1959:  $B_{1959}$**   
 (Abszisse arithmetisch skaliert)



Wir werden zunächst die Verteilungsfunktion des Einlagenbestands am Jahresende 1959 untersuchen. In Abbildung 2 ist die empirische Verteilungsfunktion von  $B_{1959}$  für den Bereich zwischen 0 DM und 2 000 DM dargestellt, d. h. es ist die kumulierte relative Häufigkeit der Sparkonten den Einlagenbeständen am Jahresende 1969 gegenübergestellt. In Analogie zu der entsprechenden Lorenzkurve (vgl. Abbildung 1) zeigt diese Darstellung, daß auf den Bereich zwischen 0 und 100 DM Einlagenbestand rund 50 % aller Sparkonten entfallen. Insofern kann eine *arithmetische* Skalierung der Abszisse den Verlauf der Verteilungsfunktion nicht voll erfassen. Wir haben daher in Abildung 3 dieselbe Verteilungsfunktion noch einmal dargestellt, allerdings mit einer *logarithmisch* skalierten Abszisse. Wenn unsere im vorangegangenen Abschnitt aufgestellte theoretische Aussage über die Verteilungsfunktion der Spareinlagenbestände zutreffen sollte, so müßte die Darstellung von Abbildung 3 den typischen Verlauf der Verteilungsfunktion einer Normalver-

teilung aufweisen. Das ist allerdings nicht der Fall, weswegen wir zunächst vermuten, daß die Verteilung der Spareinlagenbestände eine *Mischverteilung* ist, d. h. durch Überlagerung von mehreren verschiede-

*Abbildung 3*  
**Empirische Verteilungsfunktion des Spar-**  
**einlagenbestands am Jahresende 1959:  $B_{1959}$**   
 (Abszisse logarithmisch skaliert)



nen Verteilungen entsteht. Um diese Hypothese nachzuprüfen, wollen wir im folgenden die Verteilungsdichte von  $B_{1959}$  weiter analysieren, da diese insbesondere bei Mischverteilungen rein optisch und auch verfahrenstechnisch besser zu handhaben ist als die Verteilungsfunktion selbst. Die Verteilungsdichte entsteht bekanntlich aus der Verteilungsfunktion durch Differentiation, d. h. durch eine infinitesimale Betrachtungsweise. Da wir in unserem Fall eine Stichprobe, d. h. eine Realisierung einer zufälligen Größe vorliegen haben, müssen wir entsprechende *Größenklassen* bilden und durch die Verteilung der Häufigkeiten für diese Größenklassen die Verteilungsdichte approximieren. Die Größenklassen werden durch eine untere Grenze  $u_i$  und durch eine obere Grenze  $o_i$  beschrieben, wobei in die entsprechende Größenklasse diejenigen Sparkonten fallen, deren Einlagenbestand größer als  $u_i$  und kleiner oder gleich dem Wert  $o_i$  ist. Die gesamte Größenklasse wird durch das Symbol  $(u_i, o_i]$  bezeich-



net. Wenn das Spektrum des Einlagenbestands insgesamt in  $n$  Größenklassen aufgeteilt wird, so ist  $i = 1, \dots, n$ . Es ist ferner  $o_{i-1} = u_i$ .

Die letzte Größenklasse hat ebenfalls eine obere Grenze  $o_n$ , da die Anzahl der Größenklassen so gewählt wurde, daß bei der zu untersuchenden empirischen Verteilung kein Element oberhalb von  $o_n$  liegt. Die Anzahl und die Art der Größenklassen richtet sich nach der Größe des Stichprobenumfangs und nach der Form der zu untersuchenden Verteilung. Bei kleinem Stichprobenumfang darf die Größenklassenanzahl nicht zu groß gewählt werden, da sonst die Stichprobenfehler besonders stark ins Gewicht fallen. Andererseits müssen wir zur Analyse von komplexen Verteilungen (z. B. Mischverteilungen) eine möglichst große Anzahl von Größenklassen haben, da sonst die Verteilungsform nicht genau zu erfassen ist.

Die logarithmische Skalierung der Abszisse erreichen wir dadurch, indem wir die Größenklassengrenzen so bestimmen, daß der Quotient von oberer Grenze und unterer Grenze für alle Größenklassen konstant ist, d. h.

$$\frac{o_i}{u_i} = q \quad (i = 1, 2, \dots, n) .$$

In unserem speziellen Fall ergab sich durch empirische Prüfung, daß ein Faktor  $q = 1,5$  besonders sinnvoll ist. Mit diesem Faktor haben wir insgesamt 28 Größenklassen gebildet, wovon die erste Klasse Sparkonten mit Einlagenbeständen über 2 DM bis einschließlich 3 DM beinhaltet, die zweite Größenklasse zwischen 3 DM und 4,50 DM rangiert usw. Der Bereich unter 2 DM wurde nicht erfaßt, da er nur diejenigen Sparkonten enthält, die de facto erloschen sind, aber mit einer Minimaleinlage weitergeführt werden. Die letzte Größenklasse liegt zwischen rund 110 000 DM und 170 000 DM.

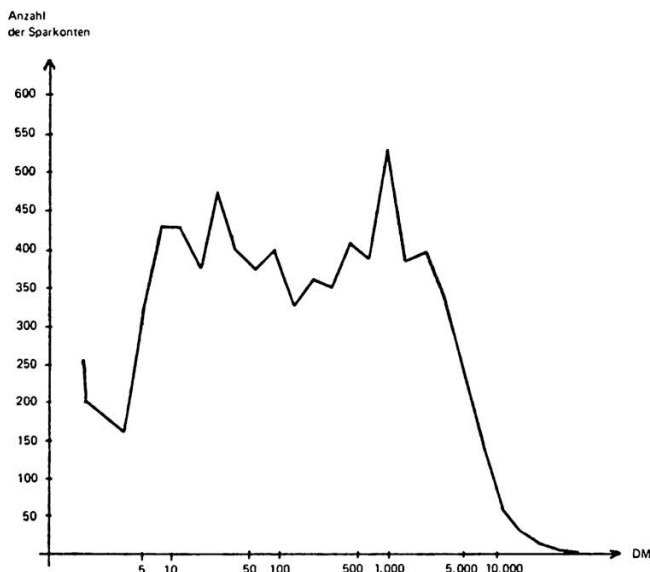
In Abbildung 4 haben wir die empirische Verteilungsdichte von  $B_{1959}$  entsprechend der oben definierten Größenklassengliederung aufgetragen. Aus dieser Darstellung kann man in groben Zügen eine Mischverteilung, die aus zwei logarithmischen Normalverteilungen besteht, herauslesen. Die relativ starken Unregelmäßigkeiten der empirischen Verteilungsdichte lassen sich durch die Stichprobe erklären.

Wir stellen demnach jetzt die Hypothese auf, daß die Verteilung der Spareinlagenbestände am Jahresende 1959, d. h. die Verteilung der Größe  $B_{1959}$  aus der Summe von zwei Lognormalverteilungen besteht.



Entsprechend können wir sagen, daß sich die Verteilung der Größe  $\log B_{1959}$  aus der Summe von zwei Normalverteilungen zusammensetzt.

*Abbildung 4*  
**Empirische Verteilungsdichte des Spar-**  
**einlagenbestands am Jahresende 1959:  $B_{1959}$**   
 (Abszisse logarithmisch skaliert)



Unter dieser Annahme sind die Abweichungen der empirischen Verteilungsdichte in Abbildung 4 von der theoretischen Verteilungsdichte nur zufällig und nur durch die Stichproben bestimmt. Unsere Aufgabe besteht nun darin, die theoretische Verteilungsdichte dadurch zu bestimmen, indem man die Verteilungsparameter so variiert, daß die empirische Verteilungsdichte optimal approximiert wird. Diesen Sachverhalt wollen wir jetzt etwas genauer erläutern:

Gemäß unserer Hypothese ist die Verteilungsdichte  $f(x)$  z. B. von  $\log B_{1959}$  eine Mischverteilung, die aus  $m$  Normalverteilungen mit den Anteilen  $a_j$ , den Mittelwerten  $\mu_j$  und den Streuungen  $\sigma_j$  mit  $j = 1, 2, \dots, m$  besteht. Dann lautet  $f(x)$  explizit:

$$(2.1) \quad f(x) = \sum_{j=1}^m \frac{1}{\sqrt{2\pi}\sigma_j} \cdot a_j \exp\left(-\frac{1}{2}(x - \mu_j)^2/\sigma_j^2\right)$$

mit  $a_j \geq 0$  und  $\sum_{j=1}^m a_j = 1$ . Ferner sei die Verteilungsdichte  $f(x)$  durch eine Stichprobe von  $n$  ungefähren Werten  $y_i$  mit

$$y_i \cong f(x_i) \quad (i = 1, 2, \dots, n)$$

ermittelt. Als Kriterium für die Güte der Anpassung zwischen theoretischer und empirischer Verteilungsdichte wählen wir die Minimum-Fehler-Quadrat-Methode, d. h. die 3  $m$  Parameter  $a_j$ ,  $\mu_j$  und  $\sigma_j$  werden aus der Gleichung

$$(2.2) \quad \sum_{i=1}^n (f(x_i) - y_i)^2 = F(a_j, \mu_j, \sigma_j)$$

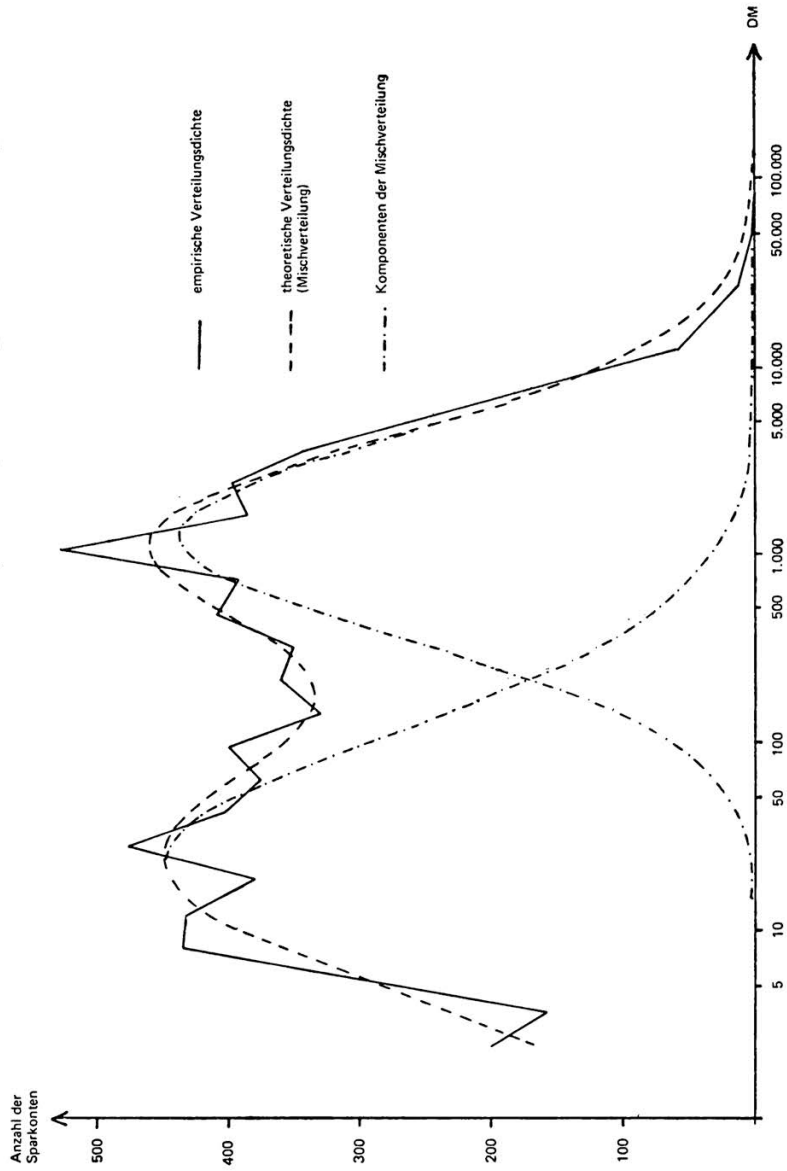
so bestimmt, daß  $F(a_j, \mu_j, \sigma_j)$  ein Minimum wird.

Eine notwendige Bedingung für diese Forderung wird bekanntlich dadurch erfüllt, daß die 3  $m$  Ableitungen dieser Funktion an der Stelle des Minimums den Wert null annehmen. Die Bestimmung der Parameter  $a_j$ ,  $\mu_j$  und  $\sigma_j$ , d. h. die Lösung des Gleichungssystems (2.2) ist numerisch relativ kompliziert, da dieses System nur transzendente Funktionen enthält. Das von uns verwandte Verfahren wollen wir hier nicht diskutieren; es sei nur erwähnt, daß es sich hierbei um ein heuristisches Gradientenverfahren handelt, das von einem beliebigen Anfangswert ausgehend im allgemeinen in wenigen Schritten konvergiert.

Die Anzahl  $m$  derjenigen Verteilungen, die gemäß Formel (2.1) die Mischverteilung  $f(x)$  bilden, ist natürlich von vornherein auch nicht bekannt. Der Verlauf der empirischen Verteilungsfunktion (Abbildung 4) läßt allerdings vermuten, daß  $m = 2$  sein wird. Diese Vermutung wurde empirisch auch dadurch bestätigt, daß bei Approximationen mit  $m = 3, 4, 5$  usw. keine wesentlichen Verteilungskomponenten mehr festgestellt werden konnten, d. h. die entsprechenden Parameter  $a_3, a_4, a_5$  usw. waren alle fast gleich null.

Mit der Hypothese, daß der Spareinlagenbestand gemäß einer Mischverteilung, die aus zwei logarithmischen Normalverteilungen besteht, verteilt ist, haben wir die Größe  $B_{1959}$  entsprechend dem oben dargestellten Verfahren analysiert und die Ergebnisse in Abbildung 5 wiedergegeben. Hier sind bei logarithmisch skaliertem Abszisse die empirische Verteilungsdichte, die hieraus ermittelte theoretische Mischverteilung und

Abbildung 5 Empirische und theoretische Verteilungsdichte und deren Komponenten für den Spareinlagenbestand am Jahresende 1959: B<sub>1959</sub> (Abszisse logarithmisch skaliert)





deren beide Komponenten dargestellt. Die aus der Abbildung ersichtliche recht gute Approximation der empirischen Verteilungsdichte durch die theoretische Dichtefunktion und deren Komponente ist allerdings keine Verifikation für die oben aufgestellte Hypothese. Hierzu müssen vielmehr statistische Hypothesentests herangezogen werden, die wir im nachfolgenden Abschnitt ausführlich diskutieren werden.

### 3. Hypothesentest und Muster der glatten Beträge

Wir wollen zunächst die oben aufgestellte Hypothese noch einmal formulieren:

Die Verteilung der Einlagenbestände auf Sparkonten am Jahresende 1959, d. h. die Verteilung der zufälligen Größe  $B_{1959}$  ist eine *Mischverteilung*, die aus *zwei Lognormalverteilungen* besteht. Diese Aussage wollen wir nun mit *Hypothesentests*, wie sie im allgemeinen in der statistischen Testtheorie benutzt werden, nachprüfen. Wir wollen allerdings vorab ausdrücklich vermerken, daß alle Hypothesentests in der Regel nur negativen Charakter haben, d. h. man kann mit solchen Tests eine Hypothese *falsifizieren*, aber *nicht verifizieren*. Bei Vorgabe einer gewissen Irrtumswahrscheinlichkeit (Fehler erster Art) kann man Kriterien aufstellen, bei deren Erfülltsein man die entsprechenden Hypothesen verwerfen muß. Die Irrtumswahrscheinlichkeit ist dann ein direktes Maß für das Verwerfen einer an sich richtigen Hypothese. Ferner sei noch erwähnt, daß man bei einem Hypothesentest im allgemeinen keine Aussagen über den sogenannten Fehler zweiter Art, d. h. über die Wahrscheinlichkeit für das Nichtverwerfen einer an sich falschen Hypothese machen kann. Es erscheint uns wesentlich, die Beschränktheit der Aussage eines Hypothesentests hier deutlich zu machen, damit aus den nachfolgenden Resultaten keine falschen Schlüsse gezogen werden. Insbesondere werden durch diese Tests über die *Richtigkeit* einer Hypothese keine Aussagen, auch nicht nach Wahrscheinlichkeit gemacht. Der bekannteste Anpassungstest für Verteilungen ist der sogenannte  $\chi^2$ -Test. Haben wir zum Beispiel die bereits oben erwähnten empirischen Werte einer Häufigkeitsverteilung  $y_i$  und die entsprechenden theoretischen Werte der Verteilungsdichte  $f(x_i)$ , so bilden wir die Prüfgröße

$$(3.1) \quad \chi^2 = N \sum_{i=1}^n (f(x_i) - y_i)^2 / f(x_i) .$$

Hierbei bedeutet  $n$  die Größenklassenanzahl der empirischen Häufigkeitsverteilung, während  $N$  den Umfang der Stichprobe angibt (in un-

serem Fall:  $n = 28$  und  $N = 12\,413$ ). Wenn nun die gemäß Formel (3.1) gebildete Prüfgröße für die Anzahl  $n$  der Größenklassen (Freiheitsgrade) bei einer vorgegebenen Irrtumswahrscheinlichkeit den entsprechenden Wert einer  $\chi^2$ -Tabelle überschreitet, so muß die zu Grunde liegende Hypothese verworfen werden. Andernfalls besteht wegen der Einseitigkeit des Hypothesentests kein Grund für das Verwerfen der Hypothese.

Ein anderer von uns im folgenden ebenfalls benutzter Hypothesentest ist von *A. Kolmogoroff* zum ersten Mal angegeben worden<sup>5</sup>. Hier wird die Prüfgröße

$$(3.2) \quad k = \max_{i=1, \dots, n} |f(x_i) - y_i|$$

gebildet, wobei die Hypothese verworfen wird, wenn der gemäß (3.2) berechnete Wert bei einer vorgegebenen Irrtumswahrscheinlichkeit größer ist als der entsprechende Zahlenwert einer Kolmogoroff-Test-Tabelle.

Prüfen wir nun die oben aufgestellte Hypothese für  $B_{1959}$  mit diesen beiden Tests, so erhalten wir für den  $\chi^2$ -Test gemäß Formel (3.1) eine Prüfgröße von  $\chi^2 = 117,9$ . Für eine Irrtumswahrscheinlichkeit von 0,1 % lautet der entsprechende Wert der  $\chi^2$ -Tabelle 52,6. Bei größeren Irrtumswahrscheinlichkeiten sind die Tabellenwerte noch kleiner. Da der berechnete Wert deutlich die Tabellenwerte überschreitet, müssen wir die oben aufgestellte Hypothese, daß für den Einlagenbestand  $B_{1959}$  eine Mischverteilung aus zwei logarithmischen Normalverteilungen vorliegt, verwerfen. Der Kolmogoroff-Test führt zum gleichen Ergebnis. Auch hier ergeben die Prüfgrößen, daß die Hypothese abgelehnt werden muß.

Trotz der optisch guten Anpassung von empirischer und theoretischer Verteilungsdichte (vergleiche Abbildung 5) führen demnach die durchgeführten statistischen Tests zu einer Ablehnung des Postulats, daß der Spareinlagenbestand  $B_{1959}$  gemäß einer lognormalen Mischverteilung verteilt ist. Ein genaues Studium von Abbildung 5 zeigt allerdings, daß die Abweichungen zwischen empirischer und theoretischer Verteilungsdichte nicht zufällig sind, sondern relativ systematischen Charakter haben. Insbesondere bei den glatten Beträgen, d. h. bei Einlagenbeständen von DM 10, DM 100, DM 500, DM 1 000 usw. zeigt die empirische Verteilungsdichte deutliche systematische Abweichungen in Form von Über-

<sup>5</sup> *A. Kolmogoroff*, Sulla determinazione empirica di una legge di distribuzione. *Giornale Istit. Ital. Attuari*, 4 (1933) S. 83.

besetzungen der entsprechenden Größenklassen, während in den dazwischen liegenden Intervallen Unterbesetzungen festzustellen sind. Diese Tatsache läßt sich dadurch erklären, daß der einzelne Sparer eine relativ starke Affinität zu glatten Beträgen, d. h. zu runden Zahlen auf seinem Sparkonto hat. Eine Analyse einer anderen Stichprobenerhebung über Einlagenbestände hat dasselbe Ergebnis gezeigt. Außerdem kann man analog zu Abbildung 5 für die Einlagenbestände in den Jahren 1960 bis 1963 die empirische und theoretische Verteilungsdichte gegenüberstellen und man erhält fast dieselben systematischen Abweichungen bei den glatten Beträgen.

Wir wollen daher aus der ursprünglichen empirischen Häufigkeitsverteilung zunächst das Muster der glatten Beträge eliminieren. Anschließend werden wir noch einmal die so bereinigte empirische Verteilungsdichte durch eine entsprechende theoretische Dichtefunktion approximieren und mittels Anpassungstests statistisch prüfen. Die Bereinigung der empirischen Verteilungsdichte vom Muster der glatten Beträge geschieht zunächst durch einen ungewichteten gleitenden Durchschnitt der Ordnung 3, d. h. zu jeder empirischen Besetzung  $y_i$  der  $i$ -ten Größenklasse berechnen wir den Wert

$$(3.3) \quad y_i^* = \frac{1}{3} (y_{i-1} + y_i + y_{i+1}) .$$

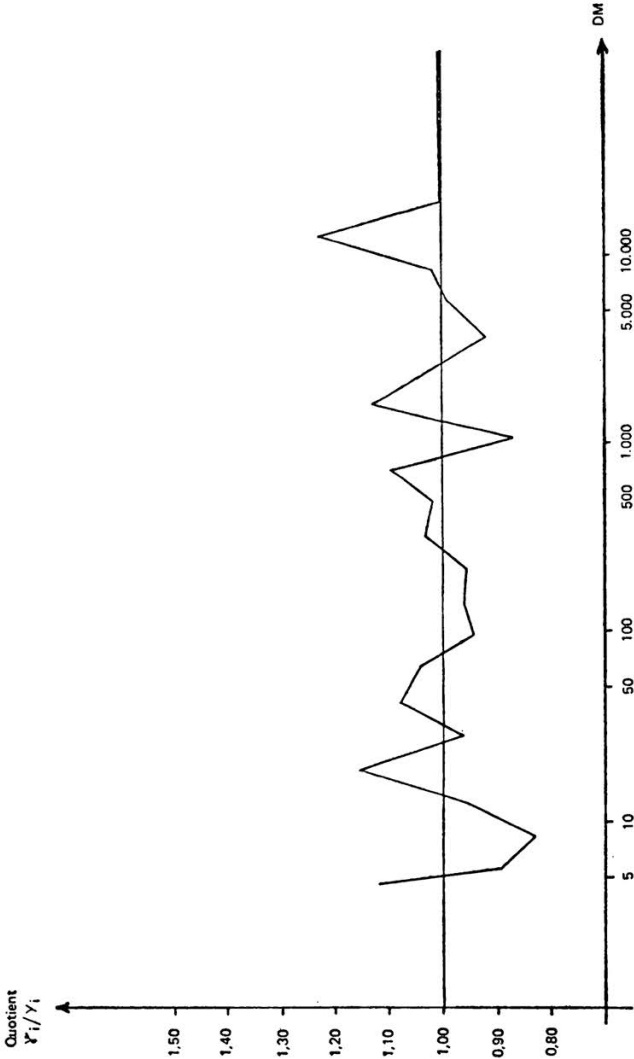
Anschließend wird eine Glättung über alle Jahre durchgeführt. Wir können nämlich für die Einlagenbestände der Jahre 1959 bis 1963 jeweils getrennt entsprechende gleitende Durchschnitte gemäß Formel (3.3) durchführen. Die sich hieraus ergebenden Werte wollen wir im folgenden mit  $y_{ik}$  bezeichnen ( $k = 1959, 1960, 1961, 1962, 1963$ ). Damit erhält man für jede Größenklassenbesetzung insgesamt fünf Werte, die den fünf Realisierungen des Einlagenbestands in den Beobachtungsjahren entsprechen. Über diese Werte wird dann noch ein gewichteter gleitender Durchschnitt der Ordnung 5 ausgeführt:

$$(3.4) \quad \gamma_{ik} = \sum_{\nu=0}^4 \alpha_{\nu} y_{i, 1959+\nu}^* / \sum_{\nu=0}^4 \alpha_{\nu} .$$

Hierbei werden die Gewichtungsfaktoren  $\alpha_{\nu}$  so gewählt, daß diejenige geglättete Größenklassenbesetzung  $y_{ik}$ , die zum  $k$ -ten Jahr gehört, das größte Gewicht erhält, während die anderen Werte  $y_{i\nu}$  entsprechend ihrem Abstand von dem jeweiligen Jahr kleinere Gewichtungsfaktoren



Abbildung 6: Muster der glatten Beträge für B<sub>1959</sub> (Abszisse logarithmisch skaliert)



erhalten. Im einzelnen werden die Gewichtungsfaktoren für das  $k$ -te Jahr wie folgt berechnet.

$$\alpha_v = 5 - |1959 + v - k| \quad (v = 0, \dots, 4; k = 1959, \dots, 1963) .$$

Die nach Formel (3.4) endgültig bereinigten Besetzungsanzahlen ergeben dann die bereinigte empirische Verteilungsdichte, die von uns anschließend weiter untersucht wird. Das Muster der glatten Beträge kann man nach dieser Bereinigung unmittelbar aus dem Quotient  $\gamma_{ik}/y_{ik}$  ablesen. Für die Größe  $B_{1959}$  ist dieses Muster in Abbildung 6 wiedergegeben. Wenn der Quotient einen Wert kleiner als 1 hat, dann war die unbereinigte empirische Verteilung an diesen Stellen ursprünglich überbesetzt. Bei einem Wert größer als 1 lag eine entsprechende Unterbesetzung vor. Die gravierenden Überbesetzungen bei den glatten Beträgen DM 10, DM 100 und DM 1 000 lassen sich in Abbildung 6 sehr deutlich erkennen. Die Muster der glatten Beträge für die Einlagenbestände in den Jahren 1960 bis 1963 sind demjenigen der Abbildung 6 so sehr ähnlich, daß sie im Rahmen der Zeichengenaugkeit nicht unterschieden werden könnten. Wir haben daher auf eine graphische Darstellung dieser Muster verzichtet. Diese starke zeitliche Konstanz des Musters rechtfertigt nochmals unsere Annahme, daß der einzelne Sparer besonders an glatten Beträgen festhält.

In Abbildung 7 haben wir die gemäß Formel (3.4) geglättete empirische Verteilungsdichte von  $B_{1959}$ , die nach der Minimum-Fehler-Quadrat-Methode ermittelte theoretische Mischverteilung und deren Komponenten bei logarithmisch skaliertter Abszisse dargestellt. Im Vergleich zu Abbildung 5 fällt auf, daß nun die empirische Verteilungsdichte einen wesentlich glatteren Verlauf hat und daher auch durch die theoretische Dichtefunktion noch besser approximiert werden kann. Ein nochmaliger Test der Hypothese, daß die bereinigte empirische Verteilung eine Mischverteilung ist, die aus zwei Lognormalverteilungen besteht, ergibt, daß sowohl beim  $\chi^2$ -Test als auch beim Kolmogoroff-Test mit relativ großen Irrtumswahrscheinlichkeiten diese Hypothese nicht verworfen werden kann. Die Aussage ist damit als statistisch gesichert zu betrachten.

Wir haben bisher die abgeleiteten Ergebnisse graphisch und auch analytisch ausschließlich für die Größe  $B_{1959}$  dargestellt. Für die Spareinlagenbestände in den anderen Jahren kommt man aber zu sehr ähnlichen Ergebnissen, auf deren graphische Wiedergabe hier verzichtet werden soll. Eine detaillierte Analyse dieser Zahlen wird im nachfolgenden Abschnitt vorgenommen.

**Abbildung 7** Geglättete empirische und theoretische Verteilungsdichte und deren Komponenten für den Spareinlagenbestand am Jahresende 1959: B<sub>1959</sub> (Abszisse logarithmisch skaliert)

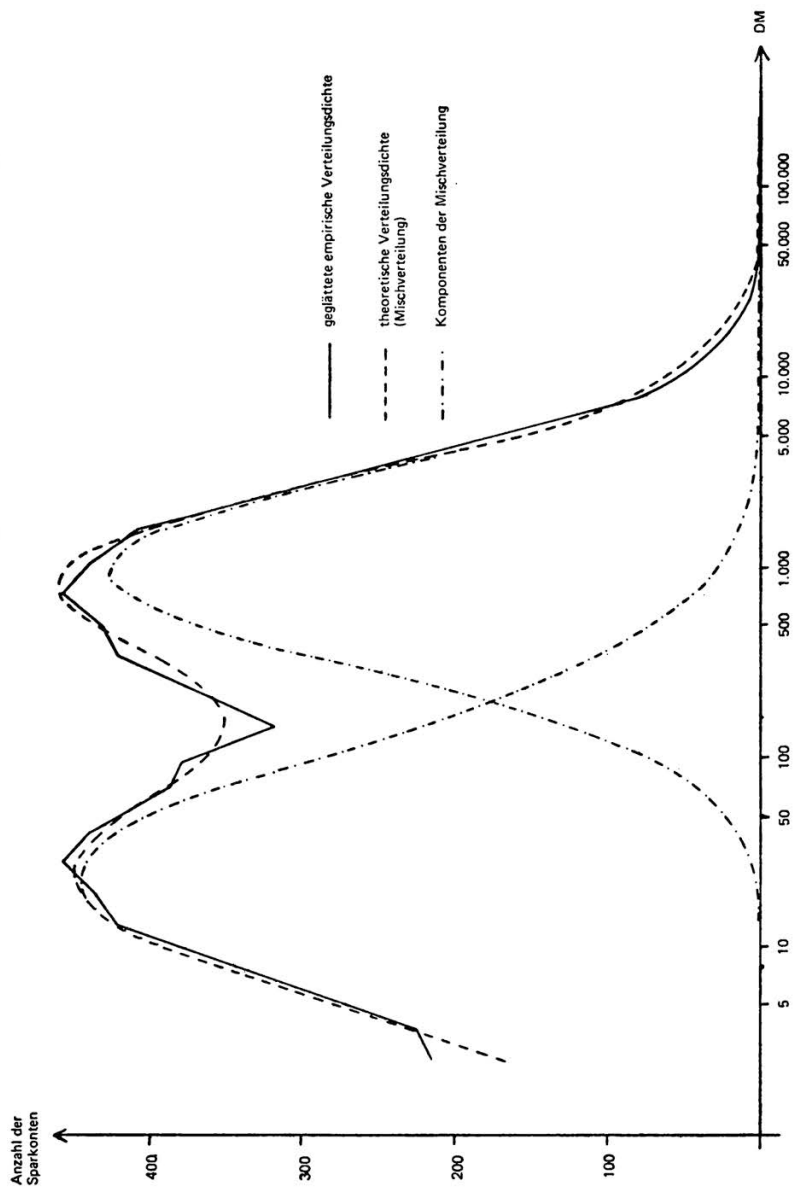




Tabelle 1

 $\chi^2$ - und Kolmogoroff-Testwerte für die Spareinlagenbestände auf Sparkonten

Größe	$\chi^2$ -Test		Kolmogoroff-Test	
	$\chi^2$ -Tabellenwerte bei 1 % Irrtumswahrscheinlichkeit	berechneter $\chi^2$ -Wert	k-Tabellenwert bei 1 % Irrtumswahrscheinlichkeit	berechneter k-Wert
B <sub>1959</sub>	44,3	12,1	0,018	0,0035
B <sub>1960</sub>	47,0	28,8	0,018	0,0026
B <sub>1961</sub>	44,3	24,2	0,017	0,0037
B <sub>1962</sub>	45,6	17,5	0,017	0,0031
B <sub>1963</sub>	47,0	18,2	0,017	0,0033

In Tabelle 1 haben wir für alle von uns untersuchten Größen den gemäß Formel (3.1) ermittelten  $\chi^2$ -Wert und den entsprechenden Tabellenwert einer  $\chi^2$ -Tabelle bei 1 % Irrtumswahrscheinlichkeit gegenübergestellt. Ebenso sind in dieser Tabelle für den Kolmogoroff-Test die gemäß Formel (3.2) berechnete Kenngröße und der zugehörige Tabellenwert für 1 % Irrtumswahrscheinlichkeit wiedergegeben. Beide Zahlenkolonnen zeigen, daß die ermittelten Kenngrößen deutlich unter den entsprechenden Tabellenwerten liegen, womit die Hypothese, daß die Einlagenbestände auf Sparkonten einer Mischverteilung aus zwei Lognormalverteilungen genügen, nicht falsifiziert werden kann.

#### 4. Zeitliche Veränderung der Verteilungsparameter

Nachdem im vorangegangenen Abschnitt lediglich die Methode zur Bestimmung der Mischverteilung und deren Komponente beispielhaft für die Größe B<sub>1959</sub> dargestellt wurde und außerdem die Hypothesentests für alle zu untersuchenden Größen nicht zur Falsifikation der entsprechenden Annahmen führten, wollen wir nun die ermittelten Verteilungsparameter darstellen und interpretieren. Wir wollen im folgenden mit *MV* die entsprechende Mischverteilung der zu untersuchenden Größen und mit *V*<sub>1</sub> bzw. *V*<sub>2</sub> deren beide Komponenten bezeichnen. Die Dichtefunktion von *MV* wollen wir mit *f*(*x*) bezeichnen, während die Dichten

von  $V_1$  bzw.  $V_2$  mit  $f_1(x)$  bzw.  $f_2(x)$  benannt werden sollen. Die Anteilsparameter für  $V_1$  und  $V_2$  seien  $a_1$  und  $a_2$ , wobei

$$(4.1) \quad a_1 + a_2 = 1, a_1, a_2 \geq 0 \text{ und } f(x) = a_1 f_1(x) + a_2 f_2(x)$$

gelten. In Tabelle 2 sind diese Anteilsparameter für die Größen  $B_{1959+i}$  mit  $i = 0, 1, \dots, 4$  wiedergegeben. Die Zahlenwerte der Tabelle zeigen eine deutliche Zunahme des Anteilsparameters  $a_1$  in den Jahren 1959 bis 1962. Das bedeutet eine kontinuierliche Zunahme der Komponente  $V_1$  relativ zur Komponente  $V_2$  und damit einer Zunahme der Gruppe derjenigen Sparer mit kleineren Einlagebeständen. Lediglich im Jahr 1963 fällt der Anteilsparameter  $a_1$  wieder zurück.

Tabelle 2

Anteilsparameter der Mischverteilungen für die Spareinlagenbestände in den Jahren 1959 bis 1963

Größe	Anteilsparameter	
	$a_1$	$a_2$
$B_{1959}$	0,57	0,43
$B_{1960}$	0,58	0,42
$B_{1961}$	0,60	0,40
$B_{1962}$	0,61	0,39
$B_{1963}$	0,56	0,44

Nun wollen wir noch die eigentlichen Verteilungsparameter interpretieren. Wie bereits oben erwähnt, sind die Verteilungen der Größen  $\log B_{1959+i}$  ( $i = 0, 1, \dots, 4$ ) Mischverteilungen, die aus zwei Normalverteilungen bestehen. Die Mittelwerte und Streuungen dieser beiden Normalverteilungen wollen wir mit  $\mu_1$  und  $\sigma_1$  bzw.  $\mu_2$  und  $\sigma_2$  bezeichnen. Dann berechnen sich die Mittelwerte der logarithmischen Normalverteilungen  $V_1$  bzw.  $V_2$ , die wir mit  $m_1$  bzw.  $m_2$  bezeichnen wollen, wie folgt:

$$(4.2) \quad m_i = e^{\mu_i + \frac{1}{2} \sigma_i^2} \quad (i = 1, 2) .$$

Die Zentralwerte  $z_1$  bzw.  $z_2$  von  $V_1$  bzw.  $V_2$  geben denjenigen kritischen Spareinlagenbestand an, der von genau 50 % aller Sparkonten,

die zu  $V_1$  bzw.  $V_2$  gehören unterschritten bzw. überschritten wird. Für diese Größen gilt

$$(4.3) \quad z_i = e^{\mu_i} \quad (i = 1, 2) .$$

Schließlich interessieren noch die Werte der größten Dichte  $g_1$  und  $g_2$ , die die jeweiligen lokalen Extrema der Verteilungsdichten  $f_1$  bzw.  $f_2$  angeben. Sie lassen sich ebenfalls aus den Parametern der entsprechenden Normalverteilungen berechnen:

$$(4.4) \quad g_i = e^{\mu_i - \sigma_i^2} \quad (i = 1, 2) .$$

Auf die Berechnungen der höheren Momente von  $V_1$  bzw.  $V_2$  wollen wir hier verzichten<sup>6</sup>.

Die obengenannten Verteilungsparameter haben wir wiederum für die Spareinlagenbestände der Jahre 1959 bis 1963 in Tabelle 3 wiedergegeben. Ein erster Blick auf diese Zahlen zeigt, daß die Dichtefunktionen der beiden Komponenten  $V_1$  und  $V_2$  sehr stark schief sind. Das ersieht man z. B. daraus, daß die Werte der größten Dichte jeweils am linken Ende der beiden Verteilungen liegen. Für  $V_1$  liegt der Wert  $g_1$  zwischen DM 1 und DM 2, während die Größe  $g_2$  zwischen DM 223 und DM 467 im Laufe der Zeit schwankt. Die eigentlichen Mittelwerte der logarithmischen Normalverteilungen  $V_1$  und  $V_2$  liegen relativ weit im rechten Teil der Verteilung. Sie steigen für beide Komponenten im Laufe der Zeit an (Ausnahme: 1963). Der Mittelwert von  $V_1$  nimmt von DM 95 auf DM 182 zu, während  $m_2$  von DM 2 246 bis DM 3 163 ansteigt. Die Zentralwerte  $z_1$  und  $z_2$  nehmen ebenfalls im Laufe der Zeit zu. Dennoch haben die Zentralwerte einen wesentlichen Abstand von den entsprechenden Mittelwerten, was ebenfalls ein Indiz für die Linksschiefe der Verteilungen ist. So ist bei  $V_1$  der entsprechende Mittelwert von  $m_1$  um mehr als das Dreifache größer als der Zentralwert  $z_1$ . Bei  $V_2$  unterscheiden sich die beiden Werte in etwa um den Faktor 2. Ferner haben wir in Tabelle 3 noch den Mittelwert der gesamten Mischverteilung, d. h. die Größe  $a_1 m_1 + a_2 m_2$  wiedergegeben. Diese Zahlenreihe stimmt relativ gut mit den aus anderen Statistiken ermittelten durchschnittlichen Einlagenbeständen je Sparkonto überein.

---

<sup>6</sup> Vgl. hierzu *J. Aitchison and J. A. C. Brown: The Lognormal Distribution with Special Reference to its Uses in Economics*. London 1956, S. 8 f.



Tabelle 3: Verteilungsparameter der Mischverteilungen für die Spareinlagenbestände in den Jahren 1959 - 1963

Größe	Komponente $V_1$					Komponente $V_2$					Mittelwert der gesamten Mischverteilung $a_1 m_1 + a_2 m_2$
	Parameter der Normalverteilung		Parameter der Lognormalverteilung			Parameter der Normalverteilung		Parameter der Lognormalverteilung			
	$\mu_1$	$\sigma_1$	Mittelwert $m_1$	Zentralwert $z_1$	Wert der größten Dichte $g_1$	$\mu_2$	$\sigma_2$	Mittelwert $m_2$	Zentralwert $z_2$	Wert der größten Dichte $g_2$	
B <sub>1959</sub>	3,19	1,65	95	24	1,58	6,95	1,24	2 246	1 040	223	1 020
B <sub>1960</sub>	3,29	1,71	116	27	1,44	7,13	1,18	2 530	1 254	308	1 129
B <sub>1961</sub>	3,31	1,73	123	28	1,38	7,29	1,17	2 891	1 459	372	1 230
B <sub>1962</sub>	3,52	1,83	182	34	1,17	7,42	1,13	3 151	1 667	467	1 341
B <sub>1963</sub>	3,45	1,73	140	31	1,58	7,29	1,24	3 163	1 460	310	1 470

### 5. Zuordnung von Sparergruppen zu den Komponenten der Mischverteilung

Wir wollen nun versuchen, die Bestimmungsgründe zu analysieren, die bedingen, daß ein Sparer oder eine Sparergruppe zur Komponente  $V_1$  bzw.  $V_2$  der Mischverteilung gehören. Bei dieser Analyse können wir allerdings nur auf diejenigen Merkmale zurückgreifen, die im Rahmen der oben erwähnten Stichprobenerhebung ermittelt wurden. Soweit andere Bestimmungsgründe für die Zugehörigkeit zu  $V_1$  bzw.  $V_2$  maßgebend sind, kann hierzu im Rahmen der nachfolgenden Untersuchung keine Aussage gemacht werden. Bezeichnen wir diejenigen Sparer, die zur Komponente  $V_1$  gehören, als Kleinsparer, so stellt sich die Frage: Welche Merkmale oder Merkmalsausprägungen bzw. Kombinationen derselben, die im Rahmen der Stichprobenerhebung ermittelt wurden, charakterisieren die Kleinsparer? Es ist evident, daß die Komponente  $V_2$  (Großsparer) durch die Komplementmenge dieser Merkmale bestimmt wird.

Zunächst wollen wir das statistische Instrumentarium für diese Analyse kurz entwickeln und darstellen; die Dichtefunktionen von  $MV$ ,  $V_1$  und  $V_2$  seien wiederum mit  $f$ ,  $f_1$  und  $f_2$  bezeichnet. Zwischen diesen Größen muß die Relation (4.1) gelten. Wenn eine zufällige Größe  $X$  z. B. zur Mischverteilungskomponente  $V_1$  gehört, d. h. gemäß der Dichtefunktion  $f_1$  verteilt ist, dann wollen wir im folgenden für diesen Sachverhalt die Abkürzungen  $L(X) = f_1$  schreiben. Analoge Abkürzungen werden für  $f$  bzw.  $f_2$  verwandt. Uns interessiert nun die bedingte Wahrscheinlichkeit, daß eine zufällige Größe  $X$  gemäß  $f_1$  verteilt ist, wenn sie den Wert  $\xi$  annimmt, d. h.

$$W(\xi) = P\{L(X) = f_1 | X = \xi\}.$$

Nach Definition gilt:

$$\begin{aligned} (5.1) \quad & P\{L(X) = f_1 | \xi \leq X < \xi + d\xi\} \\ & = P\{L(X) = f_1, \xi \leq X < \xi + d\xi\} / P\{\xi \leq X < \xi + d\xi\}. \end{aligned}$$

Ebenso gilt aber die Identität

$$\begin{aligned} (5.2) \quad & f_1(\xi) d\xi = P\{\xi \leq X < \xi + d\xi | L(X) = f_1\} \\ & = P\{L(X) = f_1, \xi \leq X < \xi + d\xi\} / P\{L(X) = f_1\}. \end{aligned}$$

Wegen  $P\{L(X) = f_1\} = a_1$

folgt aus Formel (5.2):

$$(5.3) \quad a_1 f_1(\xi) d\xi = P\{L(X) = f_1, \xi \leq X < \xi + d\xi\}.$$

Aus (5.1) und (5.3) erhalten wir dann

$$(5.4) \quad P \{L(X) = f_1 \mid \xi \leq X < \xi + d\xi\} = \frac{a_1 f_1(\xi) d\xi}{P \{\xi \leq X < \xi + d\xi\}} .$$

Schließlich ist nach Definition der Wahrscheinlichkeit und wegen (4.1):

$$(5.5) \quad P \{\xi \leq X < \xi + d\xi\} = f(\xi) d\xi = a_1 f_1(\xi) d\xi + a_2 f_2(\xi) d\xi .$$

Aus Formel (5.4) und (5.5) folgt dann für  $d\xi \rightarrow 0$ :

$$(5.6) \quad P \{L(X) = f_1 \mid X = \xi\} = \frac{a_1 f_1(\xi)}{a_1 f_1(\xi) + a_2 f_2(\xi)} .$$

Hiermit ist die obengenannte Wahrscheinlichkeit berechnet. Gemäß (5.6) erhalten wir dann eine weitere zufällige Größe

$$(5.7) \quad W = \psi(X) = \frac{a_1 f_1(X)}{a_1 f_1(X) + a_2 f_2(X)} ,$$

wobei  $X$  eine zufällige Größe aus der Mischverteilung  $MV$  ist, während für die zufällige Größe  $W$

$$(5.8) \quad 0 \leq W \leq 1$$

gilt. Nun nehmen wir aus der Gesamtheit aller Sparer eine Untergruppe heraus; diese Untergruppe habe die Mischverteilung  $MV^*$  mit der Dichte

$$(5.9) \quad \begin{aligned} f^*(\xi) &= \alpha_1 f_1(\xi) + \alpha_2 f_2(\xi) , \\ \alpha_1, \alpha_2 &\geq 0 \text{ und } \alpha_1 + \alpha_2 = 1 , \end{aligned}$$

wobei  $\alpha_1$  und  $\alpha_2$  die entsprechenden Anteilsparameter sind. Wählen wir eine zufällige Größe  $X^*$  aus  $MV^*$ , so erhalten wir gemäß (5.7) formal die zufällige Größe  $W^*$ . Deren Erwartungswert berechnet sich wie folgt:

$$(5.10) \quad E(W^*) = \int_{-\infty}^{\infty} \frac{a_1 f_1(\xi)}{a_1 f_1(\xi) + a_2 f_2(\xi)} [\alpha_1 f_1(\xi) + \alpha_2 f_2(\xi)] d\xi .$$

Setzen wir formal zur Abkürzung

$$(5.11) \quad y_1 = \int_{-\infty}^{\infty} \frac{a_1 f_1(\xi)}{a_1 f_1(\xi) + a_2 f_2(\xi)} f_1(\xi) d\xi \quad \text{und}$$

$$(5.12) \quad y_2 = \int_{-\infty}^{\infty} \frac{a_1 f_1(\xi)}{a_1 f_1(\xi) + a_2 f_2(\xi)} f_2(\xi) d\xi ,$$

so schreibt sich Formel (5.10):

$$(5.13) \quad E(W^*) = \alpha_1 y_1 + \alpha_2 y_2 .$$

Wegen (5.11) und (5.12) ist aber offensichtlich

$$(5.14) \quad a_1 = a_1 y_1 + a_2 y_2 .$$

Setzen wir nun noch wegen (5.9)  $\alpha_2 = 1 - \alpha_1$ , so erhalten wir aus (5.14) und (5.13)

$$(5.15) \quad E(W^*) = \alpha_1 y_1 + (1 - \alpha_1) (1 - y_1) \frac{a_1}{a_2} .$$

Diese Formel lösen wir nun noch nach  $\alpha_1$  auf und erhalten

$$(5.16) \quad \alpha_1 = \frac{E(W^*) - \frac{a_1}{a_2} (1 - y_1)}{y_1 - \frac{a_1}{a_2} (1 - y_1)} .$$

Mit Hilfe dieser Formel können wir demnach den Anteil  $\alpha_1$ , den eine Spareruntergruppe an der ersten Komponente  $V_1$  der Mischverteilung hat, bestimmen. Die Parameter  $a_1$ ,  $a_2$  und  $y_1$  sind bekannte Größen. Sie sind bei der Schätzung der Mischverteilung  $MV$  analytisch ermittelt worden (vgl. Abschnitt 3). Lediglich der Erwartungswert  $E(W^*)$  muß mit entsprechenden Konfidenzintervallen geschätzt werden. Hierdurch erhalten wir dann wegen (5.16) entsprechende Schätzungen mit Konfidenzintervallen für den Anteilsparameter  $\alpha_1$ .

Aus den Formeln (5.10), (5.11) und (5.12) können wir einige spezielle Werte für  $E(W^*)$  ermitteln. Es ist:

$$E(W^*) = \begin{cases} a_1 & \text{für } MV^* = MV \\ y_1 & \text{für } MV^* = V_1 \\ y_2 & \text{für } MV^* = V_2 \end{cases}$$

Wir nehmen nun an, daß wir  $N$  Stichproben  $X_i$ ,  $i = 1, \dots, N$  aus der Mischverteilung  $MV^*$  entnommen haben; dann ist wegen (5.7)

$$(5.17) \quad \overline{W} = \frac{1}{N} \sum_{i=1}^N \psi(X_i)$$



eine Schätzung für  $E(W^*)$ . Für die Streuung des Schätzwertes  $\bar{W}$  gilt bekanntlich

$$(5.18) \quad \sigma_{\bar{W}} \cong \frac{\sigma_{W^*}}{\sqrt{N-1}}.$$

Wegen (5.8) ist aber  $E(W^{*2}) \leq E(W^*)$ , so daß für die Streuung von  $W^*$  gilt

$$(5.19) \quad \sigma_{W^*}^2 = E(W^{*2}) - E(W^*)^2 \leq E(W^*) [1 - E(W^*)]$$

Durch Einsetzen von (5.19) in (5.18) erhält man

$$(5.20) \quad \sigma_{\bar{W}} \leq \frac{1}{\sqrt{N-1}} \sqrt{E(W^*) [1 - E(W^*)]} \cong \frac{1}{\sqrt{N-1}} \sqrt{\bar{W} (1 - \bar{W})}.$$

Diese Formel gestattet uns, Konfidenzgrenzen für die Schätzung  $\bar{W}$  anzugeben. Hierdurch erhalten wir wegen (5.16) auch Konfidenzgrenzen für die Schätzwerte von  $\alpha_1$ . Legen wir ein Konfidenzintervall von  $2\sigma_{\bar{W}}$ , d. h. eine statistische Sicherheit von rund 96 % für die Schätzung  $\bar{W}$  zugrunde, so erhalten wir für den Anteilsparameter  $\alpha_1$  die Konfidenzgrenzen

$$(5.21) \quad \frac{\bar{W} - \frac{2}{\sqrt{N-1}} \sqrt{\bar{W} (1 - \bar{W})} - \frac{a_1}{a_2} (1 - y_1)}{y_1 - \frac{a_1}{a_2} (1 - y_1)} \leq \alpha_1 \leq \frac{\bar{W} + \frac{2}{\sqrt{N-1}} \sqrt{\bar{W} (1 - \bar{W})} - \frac{a_1}{a_2} (1 - y_1)}{y_1 - \frac{a_1}{a_2} (1 - y_1)}.$$

Diese relativ kompliziert aussehende Formel gestattet es uns, bei Kenntnis der Anteils- und Verteilungsparameter für die Mischverteilung  $MV$  Konfidenzgrenzen für den Anteil einer Spareruntergruppe an der ersten Komponente  $V_1$  der Mischverteilung anzugeben, wenn wir aus dieser Spareruntergruppe eine Stichprobe entnehmen, mit deren Hilfe wir den Schätzwert  $\bar{W}$  bilden. Hierbei wird allerdings vorausgesetzt, daß die Spareruntergruppe ebenfalls nach einer Mischverteilung verteilt ist, deren beide Komponenten die Verteilungen  $V_1$  und  $V_2$  sind.

Wir wollen im folgenden die oben abgeleiteten theoretischen Ergebnisse anwenden und für einige interessante Spareruntergruppen den Anteil (mit Konfidenzgrenzen) an der Verteilung  $V_1$  bestimmen. Wir haben hierzu den Einlagenbestand am Jahresende 1963, d. h. die Größe  $B_{1963}$  genauer analysiert. Eine Analyse der Bestandsgrößen in den vorangegangenen Jahren führt zu relativ ähnlichen Ergebnissen, da sich die Anteilsparameter im Laufe der Zeit nur wenig ändern. Lediglich die aus Tabelle 2 für die Gesamtheit aller Sparer festzustellende Tendenz ist auch im Prinzip bei den Spareruntergruppen festzustellen. Die von uns untersuchten Spareruntergruppen sind im wesentlichen durch die Merkmale, die im Rahmen der Stichprobe ermittelt wurden, definiert. Neben einigen Merkmalen zur Person des Sparers (Geschlecht, Beruf und Geburtsjahr) sind noch weitere Analysen nach dem Wohnort, der Art des Sparkontos und der Anzahl weiterer Sparkonten desselben Sparers möglich. Die Ergebnisse der Untersuchungen sind in Tabelle 4 zusammengefaßt. Hier sind für die durch die entsprechenden Merkmalsausprägungen definierten Spareruntergruppen die gemäß Formel (5.21) berechneten Konfidenzgrenzen für den Anteil dieser Untergruppen an der Verteilung  $V_1$  wiedergegeben. Da das Konfidenzintervall symmetrisch ist, ist in einer weiteren Spalte außerdem der Mittelwert aus unterer und oberer Grenze aufgeführt. Diese Zahl ist insbesondere für eine erste Interpretation heranzuziehen.

Eine genaue Analyse der tabellierten Ergebnisse muß dem interessierten Leser überlassen bleiben. Wir wollen hier nur die wesentlichsten Ergebnisse kurz zusammenfassen: die männlichen Sparer sind in der ersten Verteilung etwas stärker vertreten als die Frauen, d. h. die Männer gehören in etwas stärkerem Maße zu den Kleinsparern. Allerdings ist dieser Unterschied nicht besonders gravierend. Stärkere Unterschiede zeigen sich schon beim Geburtsjahr. Hier sind die Sparer mit einem Geburtsjahr vor 1930 in  $V_1$  deutlich unterrepräsentiert. Sie gehören demnach eher zu den Großsparern. Hingegen gehören die sehr jungen Sparer (Geburtsjahr 1944 und später) fast ausschließlich zu den Kleinsparern. Bei einer Differenzierung nach der Wohnortgröße zeigt sich, daß die Sparer in kleinen Wohngemeinden relativ stärker zu  $V_1$  gehören und demnach auch kleinere Einlagenbestände haben. Für die regionale Differenzierung seien nur zwei besonders auffällige Tatsachen erwähnt: die Sparer im Saarland sind besonders stark in  $V_1$  vertreten, während die Sparer in Berlin überwiegend in  $V_2$  vertreten sind. Ferner können wir aus Tabelle 4 ablesen, daß nichterwerbstätige Sparer in größerem Maße über

Tabelle 4

Konfidenzgrenzen des Anteils an der Verteilungskomponente  $V_1$   
des Spareinlagenbestands  $B_{1983}$  für verschiedene Sparereruntergruppen

Spareruntergruppe	Konfidenzintervall des Anteilparameters $\alpha_1$		Mittelwert der unteren und oberen Grenze
	untere Grenze	obere Grenze	
Geschlecht des Sparers			
männlich .....	0,55	0,59	0,57
weiblich .....	0,52	0,56	0,54
Geburtsjahr des Sparers			
vor 1930 .....	0,38	0,42	0,40
1930—1943 .....	0,54	0,60	0,57
1944 und später .....	0,87	0,91	0,89
Wohnortgrößenklasse			
unter 5 000 Einwohner .....	0,59	0,64	0,61
5 000—100 000 Einwohner .....	0,51	0,56	0,53
über 100 000 Einwohner .....	0,48	0,53	0,50
Bundesland bzw. Region des Sparers- wohnorts			
Schleswig-Holstein .....	0,49	0,64	0,57
Hamburg und Bremen .....	0,44	0,56	0,50
Niedersachsen .....	0,52	0,60	0,56
Westfalen .....	0,53	0,61	0,57
Nordrhein .....	0,49	0,56	0,52
Rheinland-Pfalz .....	0,51	0,63	0,57
Hessen .....	0,48	0,57	0,52
Baden .....	0,52	0,64	0,58
Württemberg .....	0,48	0,57	0,52
Bayern .....	0,53	0,61	0,57
Saarland .....	0,49	0,73	0,61
Berlin (West) .....	0,27	0,46	0,37
Stellung des Sparers im Leben			
erwerbstätig .....	0,48	0,52	0,50
nichterwerbstätig .....	0,57	0,61	0,59

Tabelle 4 (Fortsetzung)

Spareruntergruppe	Konfidenzintervall des Anteilparameters $\alpha_1$		Mittelwert der unteren und oberen Grenze
	untere Grenze	obere Grenze	
Beruf des Sparers			
Lehrling .....	0,73	0,95	0,84
Arbeiter .....	0,64	0,69	0,66
Angestellter .....	0,51	0,57	0,54
Beamter .....	0,47	0,58	0,52
Selbständiger in Handel, Handwerk oder Gewerbe, Landwirt oder freier Beruf .....	0,50	0,57	0,54
ohne Beruf, Rentner oder Pensionär ..	0,30	0,37	0,34
Art des Sparkontos			
Sparkonto mit gesetzl. Kündigung ....	0,59	0,62	0,60
Sparkonto mit vereinb. Kündigung ..	0,38	0,48	0,43
prämienbegünstigtes Sparkonto .....	6,2	14,6	10,4
Jahr der Kontoeröffnung			
vor 1949 .....	0,61	0,67	0,64
1949—1958 .....	0,52	0,57	0,55
1959 und später .....	0,46	0,51	0,49
besitzt der Sparer weitere Sparkonten			
ja .....	0,23	0,29	0,26
nein .....	0,61	0,64	0,62
Anzahl der weiteren Sparkonten			
ein weiteres Sparkonto .....	0,28	0,35	0,32
zwei weitere Sparkonten .....	0,11	0,23	0,17
drei und mehr weitere Sparkonten ...	0,07	0,22	0,15



kleinere Spareinlagenbestände verfügen als erwerbstätige Sparer. Bei einer Aufgliederung nach dem Beruf des Sparers fällt auf, daß die Lehrlinge und Arbeiter unter den Kleinsparern überrepräsentiert sind. Eine Analyse nach dem Jahr der Kontoeröffnung zeigt, daß insbesondere die relativ alten Sparkonten in der Verteilung  $V_1$  anzutreffen sind — eine Tatsache, die nicht besonders vordergründig zu sein scheint und die auch nicht mit dem Alter des Sparers korreliert ist. Schließlich ergibt eine Auswertung nach der Anzahl der weiteren Sparkonten, daß mit steigender Kontoanzahl der Anteil an  $V_1$  stark abnimmt. Hieraus ist zu schließen, daß insbesondere dann, wenn mehrere Sparkonten vorhanden sind, auf allen diesen Konten in der Regel größere Einlagenbestände gehalten werden.

Ein großer Teil der oben gemachten Aussagen ist insofern nicht neu, als man sie auch ohne eine differenzierte Analyse vermuten würde. Durch diese Untersuchung sind die Vermutungen unter verteilungstheoretischen Gesichtspunkten verifiziert worden, wobei eine Analyse der Spareinlagenbestände unter Benutzung der vollen Information der gesamten Verteilung wesentlich schlüssiger ist als eine entsprechende Interpretation der Durchschnittswerte, wie sie üblicherweise aufgestellt wird.

## Summary

### Forms of Distribution of Savings Account Deposits

Analyses of savings account deposits have hitherto been limited essentially to investigations of average values. Predications as to the form of distribution, i. e. as to the distribution function and its parameters, could not be derived from those analyses. On the basis of comprehensive empirical material, this study shows that the deposits on savings accounts satisfy the conditions of a mixed distribution comprising two log-normal distributions. For savings deposits this verifies a predication analogous to Gibrat's law of personal income distribution. Furthermore, the two components of the mixed distribution characterize a group of small savers and one of big savers. The investigations have shown that these groups are relatively disjunct, remain almost constant over time and manifest little mutual permeability. In addition, for certain subgroups of savers defined by social criteria the study examines in what ratio they are distributed between the two components of the mixed distribution, i. e. to what extent these subgroups belong to the small or big savers.

## Résumé

### Les formes de distribution des avoirs sur comptes d'Épargne

Jusqu'à présent, les analyses des avoirs sur comptes d'épargne s'étaient bornées pour l'essentiel à effectuer des recherches sur des moyennes. Il était impossible de tirer de ces analyses des indices sur la forme de répartition, c. à. d. sur la fonction de distribution et sur ses paramètres. Sur la base de nombreuses données empiriques, le présent article constate que les dépôts d'épargne répondent à une composition mixte faite de deux répartitions logarithmico-normale. L'on a ainsi vérifié pour les dépôts d'épargne un indice analogue à celui de la répartition des revenus personnels selon la loi de Gibrat.

Les deux composantes de la répartition mixte caractérisent en outre respectivement un groupe d'épargnants modestes et un groupe de gros épargnants. Les recherches ont démontré que ces deux groupes sont relativement disjoints, demeurent presque constants sur longue période et ne sont guère perméables l'un à l'autre. Les auteurs se penchent ensuite sur certains sous-groupes d'épargnants définis par des critères sociaux, étudiant le mode de répartition des deux composantes, c. à. d. recherchant dans quelles proportions les petits et les gros épargnants se répartissent sur ces sous-groupes.