

**AUFSICHTLICHES RAHMENKONZEPT  
FÜR BACKTESTING (RÜCKVERGLEICHE)  
BEI DER BERECHNUNG DES EIGENKAPITALBEDARFS  
ZUR UNTERLEGUNG DES MARKTRISIKOS  
MIT BANKEIGENEN MODELLEN**

**Basler Ausschuss für Bankenaufsicht**

**Januar 1996**

## **Aufsichtliches Rahmenkonzept für Backtesting (Rückvergleiche) bei der Berechnung des Eigenkapitalbedarfs zur Unterlegung des Marktrisikos mit bankeigenen Modellen**

### I. Einleitung

In diesem Dokument wird das Rahmenkonzept vorgestellt, das der Basler Ausschuss für Bankenaufsicht ("der Ausschuss") für die Einbeziehung von Backtesting (Rückvergleichen) in die Berechnung des Eigenkapitalbedarfs zur Unterlegung des Marktrisikos mit bankeigenen Modellen erarbeitet hat. Es stellt eine Ausarbeitung des Absatzes B.4j) der Änderung der Eigenkapitalvereinbarung dar, die gleichzeitig veröffentlicht wird.

Viele Banken, die für die Messung des Marktrisikos ein auf internen Modellen beruhendes Verfahren anwenden, vergleichen routinemässig die täglichen Gewinne und Verluste mit den vom Modell errechneten Risikomessgrössen, um sich so ein Urteil über die Qualität und Genauigkeit ihrer Risikomesssysteme zu bilden. Dieses sogenannte Backtesting-Verfahren hat sich in zahlreichen Banken bei der Entwicklung und Einführung der Modelle zur Risikomessung als nützlich erwiesen.

Das Backtesting als Methode zur Beurteilung der Qualität eines bankeigenen Risikomodells wird laufend weiterentwickelt. Es werden stetig neue Backtesting-Verfahren erarbeitet und in breiteren Kreisen, die sich mit Risikomanagement befassen, diskutiert. Derzeit werden unter den Banken unterschiedliche Backtesting-Verfahren verwendet, und auch die Interpretationsstandards sind nicht überall dieselben. Aktive Bemühungen um eine Verbesserung und Verfeinerung der derzeit angewandten Methoden sind im Gang; das Ziel ist, genaue und ungenaue Risikomodelle besser voneinander unterscheiden zu können.

Das zentrale Element jedes Backtesting-Verfahrens ist der Vergleich der tatsächlichen Handelsergebnisse mit den vom Modell errechneten Risikomessgrössen. Liegen diese Ergebnisse nahe genug beieinander, wirft der Rückvergleich keine Fragen bezüglich der Qualität des Risikomessungsmodells auf. In einigen Fällen treten beim Backtesting jedoch derartige Differenzen zutage, dass mit fast absoluter Gewissheit Schwachpunkte bestehen, entweder im Modell selbst oder bei den Annahmen des Backtesting. Zwischen diesen beiden Extremen liegt eine Grauzone, wo die Testergebnisse, für sich genommen, nicht schlüssig sind.

Nach Ansicht des Ausschusses bietet das Backtesting die beste Gelegenheit, auf kohärente und vielfältigen Umständen gerecht werdende Weise geeignete Anreize in das Modellverfahren einzubringen. In den Stellungnahmen zu dem im April 1995 veröffentlichten Vorschlag über die Verwendung interner Modelle wurde häufig die Notwendigkeit hervorgehoben, starke Anreize für die fortlaufende Verbesserung der bankeigenen Risikomessungsmodelle zu schaffen. Bei seinen Überlegungen, wie die Rückvergleiche besser in das Modellverfahren zur Berechnung des für die Unter-

legung von Marktrisiken nötigen Eigenkapitals eingebunden werden könnten, suchte der Ausschuss zum einen der Tatsache Rechnung zu tragen, dass sich im Bankgewerbe noch keine bestimmte Backtesting-Methode durchgesetzt hat, und zum anderen auch den Bedenken wegen des unvollkommenen Signals, das aus den Rückvergleichen hervorgeht, zu begegnen.

Der Ausschuss ist der Meinung, dass das in diesem Dokument dargelegte Rahmenkonzept einen sinnvollen Mittelweg einhält zwischen der Anerkennung möglicher Grenzen des Backtesting und der Notwendigkeit, angemessene Anreize zu schaffen. Gleichzeitig ist sich der Ausschuss bewusst, dass sich die Techniken für Risikomessung und Backtesting immer noch in Entwicklung befinden, und er will wichtige Neuerungen in diesen Bereichen auch in sein Rahmenkonzept aufnehmen.

Im folgenden wird das Rahmenkonzept für das Backtesting bei der Eigenkapitalberechnung anhand interner Modelle beschrieben. Ziel dieses Rahmenkonzepts ist es, systematischere Ansätze für das Backtesting und die Interpretation von dessen Ergebnissen durch die Aufsichtsbehörde zu fördern. Der nächste Abschnitt befasst sich mit dem Charakter des Backtesting selbst, der darauf folgende Abschnitt behandelt die Interpretation der Ergebnisse durch die Aufsichtsbehörde und legt die diesbezüglich im Ausschuss vereinbarten Standards dar.

## **II. Beschreibung des Rahmenkonzepts für das Backtesting**

Das vom Ausschuss erarbeitete Rahmenkonzept für das Backtesting basiert auf dem Verfahren, das viele Banken anwenden, die sich interner Modelle für die Messung der Marktrisiken bedienen. Diese Backtesting-Programme bestehen meist aus einer in regelmässigen Abständen durchgeföhrten Gegenüberstellung der täglich gemessenen Werte des Risikopotentials der Bank ("value at risk", VAR) und des späteren tatsächlichen Tagesgewinns oder -verlusts ("Handelergebnis"). Die VAR-Messgrössen sollten bis auf einen bestimmten Bruchteil höher ausfallen als die Handelsergebnisse; der Bruchteil wird vom Konfidenzniveau der VAR-Messgrösse bestimmt. Diese Gegenüberstellung bedeutet ganz einfach, dass die Bank zählt, wie oft die Risikomessgrössen höher waren als das Handelsergebnis. Diese Zahl kann dann mit dem vorgegebenen Konfidenzniveau verglichen werden, und daraus ergibt sich die Erfolgsquote des Risikomodells der Bank. In einigen Fällen ist dieser letzte Schritt relativ informell, es können aber auch verschiedene statistische Testverfahren angewandt werden.

Das hier beschriebene Backtesting-Aufsichtskonzept beinhaltet alle im vorigen Absatz festgehaltenen Schritte, und es soll eine möglichst einheitliche Interpretation jedes einzelnen Schrittes erlauben, ohne unnötigen Aufwand zu verursachen. In dem VAR-Verfahren ist die Risikomessgrösse eine Schätzung des Betrags, der bei einer Reihe von Positionen infolge allgemeiner Marktschwankungen während einer bestimmten Haltedauer verlorengehen könnte; gemessen wird sie auf der Basis eines vorgegebenen Konfidenzniveaus.

Bei den anzuwendenden Rückvergleichen wird geprüft, ob der beobachtete prozentuale Anteil der durch die Risikomessung tatsächlich erfassten Ergebnisse einem *Konfidenzniveau von 99 %* entspricht. Es soll also ermittelt werden, ob die Risikomessgrößen des 99. Perzentils tatsächlich 99 % der Handelsergebnisse einer Bank abdecken. Zwar könnte man vorbringen, dass angesichts des Extremwert-Charakters des 99. Perzentils eine zuverlässige Schätzung schwieriger ist als bei einem niedrigeren Perzentil, doch hält es der Ausschuss für wichtig, das Testverfahren eng an das Konfidenzniveau anzulehnen, das in der Änderung der Eigenkapitalvereinbarung vorgegeben wird.

Bei der Bestimmung zweckmässiger Risikomessgrößen und Handelsergebnisse für das Backtesting ist ausserdem zu beachten, dass der VAR-Ansatz der Risikomessung im allgemeinen auf der Sensitivität eines statischen Portefeuilles gegenüber plötzlichen Preisschocks beruht. Es werden somit die Tagesendpositionen in das Risikomodell eingegeben, und dieses bewertet die mögliche Wertveränderung dieses statischen Portefeuilles infolge von Kurs-, Preis- oder Zinsänderungen während der angenommenen Haltedauer.

Theoretisch ist dies zwar einfach, in der Praxis kompliziert es jedoch das Backtesting. Beispielsweise wird oft argumentiert, dass die VAR-Messgrößen nicht mit den tatsächlichen Handelsergebnissen vergleichbar seien, da die tatsächlichen Ergebnisse zwangsläufig durch Änderungen der Zusammensetzung des Portefeuilles während der Haltedauer verzerrt werden. Nach dieser Auffassung sollten Provisionseinnahmen sowie Handelsgewinne und -verluste infolge veränderter Zusammensetzung des Portefeuilles nicht in die Definition des Handelsergebnisses einfließen, da sie nichts mit dem Risiko des statischen Portefeuilles zu tun haben, von dem bei der Bestimmung der VAR-Messgrösse ausgegangen wurde.

Dieses Argument überzeugt bezüglich der Verwendung von VAR-Messgrößen auf der Basis von Preisschocks, die auf eine längere Haltedauer ausgelegt sind. Das heisst, es wäre vermutlich nicht sehr sinnvoll, die auf 10 Tage und ein Konfidenzniveau von 99 % ausgelegten Messgrößen der Eigenkapitalberechnung mittels eines internen Modells den tatsächlichen Handelsergebnissen von 10 Tagen gegenüberzustellen. Bei stark im Handel engagierten Instituten sind erhebliche Änderungen der Portefeuille-Zusammensetzung im Vergleich zu den ursprünglichen Positionen in jedem beliebigen 10-Tage-Zeitraum die Regel. Aus diesem Grund geht das hier beschriebene Backtesting-Konzept von Risikomessgrößen aus, die auf eine Haltedauer von einem Tag ausgelegt sind. Abgesehen von den in diesem Papier genannten Einschränkungen baut das Backtesting-Verfahren auf den internen Risikomodellen der Banken auf.

Wenn man 1tägige Risikomessgrößen verwendet, sollten im Backtesting-Programm auch 1tägige Handelsergebnisse als Referenzgrösse eingesetzt werden. Allerdings gelten selbst bei 1tägigen Handelsergebnissen die oben geäusserten Bedenken über die Verzerrung der Handelergebnisse, d.h. auch diese Ergebnisse könnten für einen Vergleich ungeeignet sein, da sich darin die Wirkungen des Intraday-Handels niederschlagen, vielleicht sogar einschliesslich der Provisionserträge im Zusammenhang mit dem Verkauf neuer Produkte.

Der Handel im Tagesverlauf vergrössert tendenziell die Volatilität der Handelsergebnisse und kann auch zur Folge haben, dass das Handelsergebnis insgesamt höher ist als die

Risikomessgrösse. Dies bedeutet natürlich nicht, dass die Berechnungsmethoden für die Risikomessgrösse falsch wären, sondern ein solcher Fall liegt lediglich ausserhalb des Rahmens dessen, was mit der VAR-Methode erfasst werden soll. Die Einbeziehung von Provisionserträgen kann den Rückvergleich ebenfalls verzerrn, allerdings in die andere Richtung, da die Provisionserträge oft annuitätenähnliche Merkmale aufweisen. Da diese Erträge in der Regel nicht in die Berechnung der Risikomessgrösse einfließen, könnten durch ihre Einbeziehung in das für die Rückvergleiche verwendete Handelsergebnis Probleme des Risikomodells verdeckt werden.

Einige argumentieren, dass die von der Bank verzeichneten Handelsergebnisse die wichtigsten Zahlen für das Risikomanagement seien und dass sich die Risikomessgrössen an dieser Realität zu messen hätten, selbst wenn die den Berechnungen zugrundeliegenden Annahmen diesbezüglich Grenzen aufweisen. Andere bringen ausserdem vor, die Frage der Provisionserträge könne in ausreichender, wenn auch sehr simpler Weise dadurch gelöst werden, dass einfach der Durchschnitt der Handelsergebnisse aus der Zeitreihe herausgenommen wird, bevor die Rückvergleiche durchgeführt werden. Ein etwas raffinierteres Vorgehen bestünde darin, die Einnahmen nach Herkunft aufzuschlüsseln (z.B. Provisionen, Spreads, Marktveränderungen, Intraday-Handelsergebnisse).

Soweit das Backtesting-Programm als rein statistische Überprüfung der Genauigkeit der VAR-Berechnung angesehen wird, ist es eindeutig am zweckmässigsten, das tägliche Handelsergebnis so zu definieren, dass ein Test ohne Verzerrungen möglich ist. Um dieser Forderung zu entsprechen, sollten die Banken sich die Möglichkeit schaffen, Rückvergleiche durchzuführen, die auf den hypothetischen Änderungen des Portefeuillewerts bei unveränderten Tagesendpositionen beruhen.

Auch ein Backtesting anhand der tatsächlichen Tagesgewinne und -verluste ist nützlich, da es Fälle aufdecken kann, in denen die Risikomessgrössen die Handelsvolatilität nicht richtig erfassen, obwohl sie auf korrekte Weise berechnet werden.

Aus diesen Gründen legt es der Ausschuss den Banken nahe, die notwendigen Vorkehrungen zu treffen, um Backtesting sowohl anhand hypothetischer als auch anhand tatsächlicher Handelsergebnisse durchführen zu können. Auch wenn die nationalen Aufsichtsbehörden diese verschiedenen Ansätze vielleicht unterschiedlich gewichten, hat eindeutig jeder seine Vorzüge. Werden die beiden Ansätze kombiniert, dürften sie ein gutes Bild vom Verhältnis zwischen den berechneten Risikomessgrössen und den Handelsergebnissen vermitteln.

Der nächste Schritt in der Spezifizierung des Backtesting-Programms betrifft die Art des Rückvergleichs selbst und die Häufigkeit, mit der er durchgeführt werden soll. Das vom Ausschuss erarbeitete Konzept ist das einfachste Verfahren für die Gegenüberstellung der Risikomessgrössen und der Handelsergebnisse und besteht darin, dass berechnet wird, wie oft die Handelsergebnisse nicht von der Risikomessung erfasst werden ("Ausnahmen"). Zum Beispiel sollte bei einem Konfidenzniveau von 99 % in einem Zeitraum von 200 Handelstagen die tägliche Risikomessgrösse im Durchschnitt 198 der 200 Handelsergebnisse erfassen, so dass zwei Ausnahmen übrigbleiben.

Was die Häufigkeit der Rückvergleiche betrifft, so ist der Wunsch, die Rückvergleiche auf so viele Werte wie möglich abzustützen, gegen den Wunsch abzuwägen, sie regelmässig durch-

zuführen. Das anzuwendende Backtesting-Verfahren beinhaltet ein *formelles Prüfen und Dokumentieren von Ausnahmen auf Quartalsbasis*, wobei die Daten der letzten zwölf Monate einzusetzen sind.

Die Umsetzung des Backtesting-Programms sollte formell an dem Tag beginnen, an dem die Eigenkapitalvorschriften für die Berechnung anhand interner Modelle in Kraft treten, d.h. spätestens zum Jahresende 1997. Danach würde *das erste formelle Dokumentieren von Ausnahmen im Rahmen des Backtesting-Programms spätestens Ende 1998 erfolgen*. Das schliesst natürlich nicht aus, dass nationale Aufsichtsbehörden schon vorher Backtesting-Ergebnisse anfordern können, und insbesondere können diese Ergebnisse - nach Ermessen der nationalen Aufsichtsbehörde - auch bereits in das Genehmigungsverfahren für die bankeigenen Modelle einbezogen werden.

Die Verwendung der Daten der letzten zwölf Monate ergibt rund 250 Tageswerte für die Rückvergleiche. *Die nationale Aufsichtsbehörde stützt sich für ihr weiteres Vorgehen auf die Anzahl der gegenüber dem Modell der Bank festgestellten Ausnahmen (von den 250 Werten)*. In vielen Fällen wird es kein weiteres Vorgehen geben. In anderen Fällen leitet die Aufsichtsbehörde vielleicht Besprechungen mit der Bank ein, um festzustellen, ob im Modell der Bank ein Problem vorliegt. In den gravierendsten Fällen kann die Aufsichtsbehörde die Eigenkapitalanforderung der Bank heraufsetzen oder die Verwendung des Modells verbieten.

Die Verwendung der Zahl der Ausnahmefälle als vorrangige Referenzgrösse im Backtesting-Verfahren bietet sich an, weil dieser Ansatz einfach und unkompliziert ist. Aus der Sicht des Statistikers erfordert die Verwendung der Zahl der Ausnahmen als Grundlage für die Beurteilung des Modells einer Bank relativ wenige feste Annahmen. Die wichtigste Annahme ist, dass das Testergebnis jedes Tages (Ausnahme/keine Ausnahme) vom Ergebnis jedes anderen Tages unabhängig ist.

Dem Ausschuss ist natürlich klar, dass Tests dieser Art nur in begrenztem Umfang zwischen einem genauen und einem ungenauen Modell unterscheiden können. Für einen Statistiker heisst das, dass es unmöglich ist, den Test so auszurichten, dass er korrekt alle problematischen Modelle signalisiert, ohne Fehlalarm in bezug auf viele andere zu geben. Diese Begrenzung spielte bei der Gestaltung des hier präsentierten Konzepts eine wesentliche Rolle; sie sollte auch von den nationalen Aufsichtsbehörden bei der Interpretation der Ergebnisse des Backtesting-Programms einer Bank gebührend beachtet werden. Der Ausschuss sieht diese Begrenzung aber nicht als entscheidenden Einwand gegen das Backtesting an, sondern er hält es für besser, die aufsichtlichen Standards auf ein klares Rahmenkonzept abzustützen, selbst wenn dieses seine Grenzen und Mängel hat, als nach reinem Ermessen vorzugehen oder einen Standard ohne jeglichen Anreiz zu schaffen.

### **III. Rahmenkonzept für die Interpretation der Backtesting-Ergebnisse durch die Aufsichtsbehörden**

#### **a) Das Drei-Zonen-Konzept**

Der Ausschuss dachte bei der Entwicklung des Rahmenkonzepts für die Interpretation der Backtesting-Ergebnisse durch die Aufsichtsbehörden an die statistischen Grenzen der Rückvergleiche; das Rahmenkonzept enthält deshalb eine Reihe von möglichen Reaktionen, je nach der Stärke des Signals, das aus dem Backtesting hervorgeht. Diese Reaktionen werden in drei abgestufte, mit Farben bezeichnete Zonen eingeteilt. Die "grüne" Zone entspricht Ergebnissen, die an und für sich nicht auf ein Problem bezüglich der Qualität oder der Genauigkeit des Modells einer Bank hinweisen. In die "gelbe" Zone fallen Ergebnisse, die diesbezüglich zwar Fragen aufwerfen, aber keinen endgültigen Schluss zulassen. Ein Backtesting-Ergebnis in der "roten" Zone deutet mit fast völliger Sicherheit auf ein Problem im Risikomodell einer Bank hin.

Der Ausschuss hat bei der Definition dieser Zonen Standards bezüglich der Anzahl der Ausnahmen vereinbart, die sich beim Backtesting ergeben. Die Standards werden weiter unten dargestellt. Um diese Definitionen jedoch in die richtige Perspektive zu rücken, soll zunächst untersucht werden, mit welcher Wahrscheinlichkeit unterschiedliche Annahmen bezüglich der Genauigkeit des Risikomodells einer Bank jeweils zu einer unterschiedlichen Zahl von Ausnahmen führen.

#### **b) Statistische Erwägungen bei der Definition der Zonen**

Es wurden drei Zonen festgelegt, und sie wurden so abgegrenzt, dass damit möglichst ein Mittelweg zwischen zwei Arten von statistischen Fehlern eingeschlagen wird: 1) dass ein genaues Risikomodell aufgrund des Backtesting-Ergebnisses als ungenau eingestuft wird; 2) dass ein ungenaues Modell aufgrund des Backtesting-Ergebnisses nicht als solches erkannt wird.

Tabelle 1 zeigt die Wahrscheinlichkeit auf, mit der sich bei einer Stichprobe von 250 voneinander unabhängigen Werten eine bestimmte Zahl von Ausnahmen ergibt, wobei von mehreren Annahmen über den tatsächlichen Prozentsatz der vom Modell erfassten Ergebnisse ausgegangen wird (d.h. es handelt sich um Binomialwahrscheinlichkeiten). Zum Beispiel sind im linken Teil von Tabelle 1 die mit einem genauen Modell verbundenen Wahrscheinlichkeiten aufgeführt (d.h. hier besteht ein tatsächliches Erfassungsniveau von 99 %). Bei diesen Annahmen geht aus der mit "exakt" bezeichneten Spalte hervor, dass in 6,7 % der Stichproben mit exakt 5 Ausnahmen zu rechnen ist.

Im rechten Teil von Tabelle 1 sind die mit mehreren möglicherweise ungenauen Modellen verbundenen Wahrscheinlichkeiten aufgeführt, d.h. mit Modellen, deren tatsächliches Erfassungsniveau 98 %, 97 %, 96 % oder 95 % beträgt. Aus der mit "exakt" bezeichneten Spalte geht so bei einer angenommenen Erfassung von 97 % hervor, dass in 10,9 % der Stichproben mit 5 Ausnahmen zu rechnen ist.

Tabelle 1 zeigt überdies mehrere wichtige Fehlerwahrscheinlichkeiten an. Bei der Annahme, dass das Modell 99 % der Ergebnisse erfasst (das gewünschte Erfassungsniveau), geht aus der Tabelle hervor, wie wahrscheinlich es ist, dass die Festlegung einer bestimmten Zahl von Ausnahmen als Schwelle für die Rückweisung eines Modells zur irrtümlichen Ablehnung eines genauen Modells führt (Fehler des Typs 1). Wird die Schwelle z.B. schon bei 1 Ausnahme angesetzt, so werden in 91,9 % der Fälle genaue Modelle zurückgewiesen, da sie nur in den 8,1 % der Fälle, wenn sie null Ausnahmen erzeugen, der Rückweisung entgehen. Je höher die Schwelle der Ausnahmen angesetzt wird, desto mehr nimmt die Wahrscheinlichkeit dieses Fehlers ab.

Bei den Annahmen, dass das tatsächliche Erfassungsniveau des Modells unter 99 % liegt, geht aus der Tabelle hervor, mit welcher Wahrscheinlichkeit die Festlegung einer bestimmten Zahl von Ausnahmen als Schwelle für die Rückweisung eines Modells zur irrtümlichen Zulassung eines Modells mit dem angenommenen (ungenauen) Erfassungsniveau führt (Fehler des Typs 2). Wenn beispielsweise das tatsächliche Erfassungsniveau eines Modells 97 % beträgt und die Schwelle für die Rückweisung bei 7 oder mehr Ausnahmen angesetzt wird, gibt die Tabelle an, dass dieses Modell in 37,5 % der Fälle irrtümlich zugelassen würde.

Bei der Interpretation der Daten von Tabelle 1 ist ferner zu beachten, dass die Alternativmodelle zwar bezüglich der Wahrscheinlichkeit nahe bei dem gewünschten Standard zu liegen scheinen (97 % liegt nahe bei 99%), dass jedoch der Unterschied bezüglich der errechneten Risikomessgrößen beträchtlich sein kann, d.h. die Risikomessgröße einer Bank kann zwar erheblich unter der eines genauen Modells liegen, aber immer noch 97 % der Handelsergebnisse abdecken. Bei normalverteilten Handelsergebnissen entspricht das 97. Perzentil z.B. 1,88 Standardabweichungen, während das 99. Perzentil 2,33 Standardabweichungen entspricht, also fast 25 % mehr. Es ist daher sehr begreiflich, dass die Aufsichtsbehörden zwischen Modellen mit einer Erfassungsquote von 99 % und Modellen mit einer Quote von z.B. 97 % unterscheiden wollen.

### c) ***Definition der grünen, der gelben und der roten Zone***

Die Resultate in Tabelle 1 zeigen auch einige der statistischen Grenzen des Backtesting auf. Insbesondere gibt es keine bestimmte Zahl von Ausnahmen, bei denen sowohl die Wahrscheinlichkeit, irrtümlicherweise ein genaues Modell zurückzuweisen, als auch die Wahrscheinlichkeit, irrtümlicherweise alle betreffenden ungenauen Modelle zuzulassen, gering ist. Aus diesem Grund hat sich der Ausschuss gegen einen Ansatz mit einer einzigen Schwellenzahl entschieden.

In Erkenntnis dieser Grenzen hat der Ausschuss die Ergebnisse in drei Kategorien eingeteilt. In der ersten Kategorie entsprechen die Testergebnisse einem genauen Modell, und die Wahrscheinlichkeit, irrtümlicherweise ein ungenaues Modell zuzulassen, ist gering (grüne Zone). Am entgegengesetzten Ende ist es höchst unwahrscheinlich, dass die Testergebnisse mit einem genauen Modell erzielt wurden, und die Wahrscheinlichkeit, dass ein genaues Modell auf dieser Basis zurückgewiesen wird, ist sehr gering (rote Zone). Zwischen diesen beiden Extremen liegt jedoch eine Zone, in der die Backtesting-Ergebnisse sowohl aus einem genauen als auch aus einem ungenauen Modell

hervorgegangen sein könnten; in diesem Fall sollte die Aufsichtsbehörde, ehe sie Massnahmen ergreift, von der betreffenden Bank zusätzliche Informationen über ihr Modell verlangen (gelbe Zone).

In Tabelle 2 sind die vom Ausschuss vereinbarten Abgrenzungen für diese Zonen dargestellt sowie die wahrscheinliche Reaktion der Aufsichtsbehörde für jedes Backtesting-Ergebnis, gestützt auf eine Stichprobe von 250 Werten. Bei einer anderen Stichprobengröße sind die Abgrenzungen durch Berechnung der Binomialwahrscheinlichkeiten bei einer tatsächlichen Erfassung von 99 % zu ermitteln (s. Tabelle 1). Die gelbe Zone beginnt an dem Punkt, an dem die Wahrscheinlichkeit, die angegebene oder eine niedrigere Zahl von Ausnahmen zu erhalten, gleich oder grösser als 95 % ist. Tabelle 2 zeigt diese kumulativen Wahrscheinlichkeiten für jede Zahl von Ausnahmen. Bei 250 Werten ist z.B. ersichtlich, dass bei einer tatsächlichen Erfassungsquote von 99 % in 95,88 % der Fälle 5 oder weniger Ausnahmen erreicht werden. Die gelbe Zone beginnt somit bei 5 Ausnahmen.

Gleicherweise ist der Anfang der roten Zone als der Punkt definiert, an dem die Wahrscheinlichkeit, die angegebene oder eine niedrigere Zahl von Ausnahmen zu erhalten, gleich oder grösser als 99,99 % ist. Tabelle 2 zeigt, dass bei einer Stichprobe von 250 Werten und einer tatsächlichen Erfassungsquote von 99 % dieser Punkt bei 10 Ausnahmen liegt.

#### **d) Die grüne Zone**

Die grüne Zone bedarf nur weniger Erläuterungen. Da es bei einem Modell mit einer tatsächlichen Erfassungsquote von 99 % durchaus wahrscheinlich ist, dass in einer Stichprobe von 250 Werten bis zu 4 Ausnahmen auftreten, besteht kaum Anlass zu Sorge, wenn die Backtesting-Ergebnisse in dieser Bandbreite liegen. Hierfür sprechen auch die Daten von Tabelle 1; danach ist die Gefahr, irrtümlicherweise ein ungenaues Modell zuzulassen, gering, wenn Ergebnisse in dieser Spanne akzeptiert werden.

#### **e) Die gelbe Zone**

Die Spanne von 5 bis 9 Ausnahmen bildet die gelbe Zone. Ergebnisse in dieser Bandbreite sind bei genauen und bei ungenauen Modellen plausibel, Tabelle 1 deutet allerdings darauf hin, dass sie bei letzteren eher auftreten. Ferner ist aus Tabelle 1 ersichtlich, dass die Vermutung, dass das Modell ungenau ist, um so stichhaltiger wird, je höher die Zahl der Ausnahmen innerhalb der Spanne von 5 bis 9 ist.

Der Ausschuss ist übereingekommen, dass in der gelben Zone die Zahl der Ausnahmen grundsätzlich als Richtschnur dafür dienen sollte, in welchem Umfang die Aufsichtsbehörde die Eigenkapitalanforderungen einer Bank ggf. erhöht. Tabelle 2 gibt die für Backtesting-Ergebnisse in der gelben Zone vereinbarten Richtlinien des Ausschusses für eine Erhöhung der Multiplikationsfaktoren an, die für die Eigenkapitalberechnung mittels interner Modelle gelten.

Diese Richtlinien tragen dazu bei, die erwünschte Anreizstruktur bei der Verwendung interner Modelle zu wahren. Insbesondere erhöht sich der von der Aufsichtsbehörde ggf. verhängte

Aufschlag mit der Zahl der Ausnahmen. Die in Tabelle 1 aufgeführten Ergebnisse untermauern die These, dass 9 Ausnahmen mehr Anlass zu Beunruhigung geben als 5; dies soll in diesen Schritten zum Ausdruck kommen.

Den gewählten Werten liegt der generelle Gedanke zugrunde, dass die Erhöhung des Multiplikationsfaktors ausreichen sollte, um das Modell wieder auf den Standard eines Konfidenzniveaus von 99 % zu bringen. Zum Beispiel implizieren 5 Ausnahmen in einer Stichprobe von 250 Werten eine Erfassungsquote von nur 98 %. Die Erhöhung des Multiplikationsfaktors sollte somit gross genug sein, dass das Modell mit der Erfassungsquote von 98 % in eines mit einer Erfassungsquote von 99 % umgewandelt wird. Selbstverständlich sind für exakte Berechnungen dieser Art zusätzliche statistische Annahmen erforderlich, die nicht in allen Fällen standhalten dürften. Geht man z.B. von einer Normalverteilung der Handelsergebnisse aus, dann ist das Verhältnis zwischen dem 99. und dem 98. Perzentil rund 1,14, und bei einem Multiplikationsfaktor von 3 ist somit eine Erhöhung um rund 0,40 erforderlich. Wenn die tatsächliche Verteilung aber nicht normal ist, sondern sogenannte "fat tails" aufweist, dann können grössere Erhöhungen nötig sein, um den Standard des 99. Perzentils zu erreichen. Die Sorge hinsichtlich "fat tails" war im übrigen ein wichtiger Faktor bei der Festlegung der in Tabelle 2 aufgeführten Erhöhungen.

Es ist jedoch zu betonen, dass diese Erhöhungen nicht völlig automatisch erfolgen sollten. Aus Tabelle 1 ist ersichtlich, dass die Ergebnisse in der gelben Zone nicht immer ein Hinweis auf ein ungenaues Modell sind, und der Ausschuss ist nicht daran interessiert, die Banken für reines Pech zu bestrafen. *Um jedoch das Anreizsystem nicht zu beeinträchtigen, sollte grundsätzlich davon ausgegangen werden, dass Backtesting-Ergebnisse in der gelben Zone eine Erhöhung des Multiplikationsfaktors zur Folge haben, es sei denn, die Bank kann nachweisen, dass eine solche Erhöhung ungerechtfertigt ist.*

Mit anderen Worten: Die Beweislast liegt bei der Bank - nicht die Aufsichtsbehörde muss beweisen, dass ein Problem vorliegt, sondern die Bank muss beweisen, dass ihr Modell grundsätzlich solide ist. In einer solchen Situation können vielerlei zusätzliche Informationen für die Beurteilung des Modells der Bank von Bedeutung sein.

Zum Beispiel wären dann diejenigen Backtesting-Ergebnisse besonders interessant, bei denen die Handelstätigkeit der Bank in Teilbereiche aufgeschlüsselt ist. Viele Banken, die regelmässig Rückvergleiche durchführen, gliedern ihren Handelsbestand nach Risikofaktoren oder Produktkategorien auf. Dadurch können unter Umständen Probleme, die auf der Ebene des Gesamtportefeuilles auftreten, zu ihrer Quelle in einem bestimmten Handelsbereich oder Risikomodell zurückverfolgt werden.

Darüber hinaus sollten die Banken alle Ausnahmen, die sich aus ihrem aktuellen Backtesting-Programm ergeben, dokumentieren, ebenso etwaige Erklärungen für die Ausnahmen. Dies ist wichtig, wenn die Aufsichtsbehörde über angemessene Massnahmen bei einem Backtesting-Ergebnis in der gelben Zone entscheiden muss. Die Banken führen möglicherweise Rückvergleiche auch für andere Konfidenzniveaus als 99 % oder andere, hier nicht besprochene statistische Tests durch. Auch solche Informationen können sich natürlich bei der Beurteilung ihres Modells als nützlich erweisen.

In der Praxis gibt es mehrere mögliche Erklärungen für die beim Backtesting zutage tretenden Ausnahmen. Einige betreffen die grundlegende Integrität des Modells, andere lassen auf ein zu wenig spezifiziertes oder qualitativ ungenügendes Modell schliessen, wieder andere deuten entweder auf unglückliche Zufälle oder auf schlechte Ergebnisse im Intraday-Handel hin. Es kann sehr nützlich sein, die sich aus dem Modell einer Bank ergebenden Ausnahmen in diese Kategorien einzuteilen.

*Grundlegende Integrität des Modells*

1. Die Systeme der Bank erfassen ganz einfach nicht das Risiko der Positionen selbst.  
(Beispiel: Die Positionen einer Auslandsniederlassung werden unkorrekt gemeldet.)
2. Die Volatilitäten und/oder Korrelationen des Modells wurden falsch berechnet.  
(Beispiel: Der Rechner dividiert durch 250 statt - richtigerweise - durch 225.)

*Das Modell ist verbesserungsbedürftig*

3. Das Risikomodell schätzt das Risiko einiger Instrumente zu wenig genau ein.  
(Beispiel: zu wenige Laufzeitänder oder ein nicht berücksichtigter Spread.)

*Unglücklicher Zufall oder die Märkte entwickelten sich anders, als vom Modell vorhergesesehen*

4. Reiner Zufall (ein Ereignis von sehr geringer Wahrscheinlichkeit).
5. Ein Markt veränderte sich stärker, als vom Modell vorhergeschen (d.h. die Volatilität war erheblich höher als erwartet).
6. Verschiedene Märkte entwickelten sich nicht synchron wie erwartet (d.h. die Korrelationen unterschieden sich erheblich von den Annahmen des Modells).

*Intraday-Handel*

7. Zwischen dem Ende des ersten Tages (als die Risikoschätzung berechnet wurde) und dem Ende des zweiten Tages (als die Handelsergebnisse verbucht wurden) kam es zu einer massiven (und verlustbringenden) Änderung in den Positionen der Bank oder trat ein anderes ertragsrelevantes Ereignis ein.

Im allgemeinen sind Probleme im Zusammenhang mit der grundlegenden Integrität des Risikomodells potentiell die gravierendsten. Fallen Ausnahmen für eine bestimmte Handelseinheit in diese Kategorie, sollte der Aufschlagfaktor angewendet werden. Ausserdem muss das Modell vielleicht gründlich überarbeitet oder angepasst werden, und die Aufsichtsbehörde sollte dafür sorgen, dass dies geschieht.

Mit der zweiten Kategorie von Problemen (mangelnde Präzision des Modells) ist eigentlich bei den meisten Risikomodellen irgendwann zu rechnen. Man kann von keinem Modell erwarten, dass es unendlich genau ist, und jedes Modell beinhaltet somit bis zu einem gewissen Grad Näherungen. Hat eine Aufsichtsbehörde allerdings den Eindruck, das Modell einer bestimmten Bank sei für diese Art von Problemen besonders anfällig, sollte sie den Aufschlagfaktor verhängen und weitere Anreize für Verbesserungen überlegen.

Die dritte Problemkategorie (die Märkte haben sich in einer vom Modell nicht vorhergesehenen Weise entwickelt) dürfte ebenfalls zumindest zeitweise bei allen VAR-Modellen auftreten. Insbesondere wird auch von einem genauen Modell nicht erwartet, dass es 100 % der Handelsergebnisse abdeckt. Einige Ausnahmen sind sicher dem Zufallsbereich von 1 % zuzuschreiben, dessen Erfassung vom Modell gar nicht erwartet wird. In anderen Fällen kann sich das Verhalten der Märkte derart ändern, dass vorherige Schätzungen bezüglich Volatilität und Korrelationen nicht mehr adäquat sind. Kein VAR-Modell ist gegen diese Art Problem immun; dieses ist unvermeidlich, wenn früheres Marktverhalten als Grundlage für die Einschätzung des Risikos zukünftiger Marktentwicklungen dient.

Je nach der für das Backtesting angewandten Definition der Handelsergebnisse können Ausnahmen schliesslich auch durch Intraday-Handelsergebnisse oder ein ungewöhnliches Geschehnis bei den Handelerträgen, die nicht aus dem Eingehen von Positionen stammen, zustande kommen. Obwohl Ausnahmen mit solcher Ursache nicht unbedingt auf ein Problem im VAR-Modell der Bank hindeuten, können sie trotzdem der Aufsichtsbehörde Anlass zu Sorge geben; es ist daher das Verhängen des Aufschlags in Betracht zu ziehen.

Eine weitere wichtige Information ist, wie stark das Handelsergebnis die Risikomessgrösse übersteigt. Bei sonst gleichen Voraussetzungen geben sehr grosse Abweichungen mehr Anlass zu Sorge als geringfügige.

Wenn die Aufsichtsbehörde entscheiden muss, ob sie bei der Eigenkapitalberechnung einer Bank Aufschläge anwenden soll oder nicht, könnte sie diese Faktoren und auch noch andere in Erwägung ziehen; z.B. kann sie prüfen, wie weit die Bank die geltenden qualitativen Standards für das Risikomanagement einhält. Anhand der von der Bank gelieferten zusätzlichen Informationen kann die Aufsichtsbehörde dann das angemessene Vorgehen festlegen.

Im allgemeinen ist das Verhängen einer höheren Eigenkapitalanforderung für Ergebnisse in der gelben Zone durchaus angemessen, wenn die Aufsichtsbehörde der Ansicht ist, dass die Ursache eines solchen Ergebnisses ein korrigierbares Problem im Modell einer Bank ist. Anders ist es bei einem unerwarteten Auftreten hoher Marktvolatilität, das vermutlich die meisten Modelle nicht vorherzusehen vermögen. Solche Ereignisse können zwar eine Belastung darstellen, aber sie zeigen nicht unbedingt an, dass das Risikomodell einer Bank neu gestaltet werden müsste. Bei ernsthaften Problemen im Zusammenhang mit der grundlegenden Integrität des Modells sollte die Aufsichtsbehörde ein Verbot des betreffenden Modells für die Eigenkapitalberechnung erwägen.

## f) Die rote Zone

Im Gegensatz zu Backtesting-Ergebnissen in der gelben Zone, bei denen die Aufsichtsbehörde bei der Interpretation einen Ermessensspielraum hat, sollten Ergebnisse in der roten Zone (10 oder mehr Ausnahmen) automatisch zu der Vermutung führen, dass das Modell einer Bank mit einem Problem behaftet ist, denn es ist höchst unwahrscheinlich, dass ein genaues Modell bei einer Stichprobe von 250 Handelsergebnissen zu 10 oder mehr voneinander unabhängigen Ausnahmen führt.

Wenn das Modell einer Bank in die rote Zone fällt, sollte die Aufsichtsbehörde daher den Multiplikationsfaktor für das betreffende Modell automatisch um 1 (von 3 auf 4) erhöhen. Selbstverständlich sollte sie auch die Gründe für die zahlreichen Fehlberechnungen des Modells untersuchen, und die Bank ist aufzufordern, unverzüglich eine Verbesserung ihres Modells in Angriff zu nehmen.

Zwar sind 10 Ausnahmen bei 250 Werten sehr viel, doch in ganz seltenen Fällen kann es triftige Gründe dafür geben, dass ein genaues Modell so viele Ausnahmen produziert. Vor allem wenn an Finanzmärkten grosse Systemänderungen im Gang sind, ist damit zu rechnen, dass sich auch viele Volatilitäten und Korrelationen verändern, möglicherweise sogar beträchtlich. Wenn eine Bank dann nicht in der Lage ist, ihre Volatilitäts- und Korrelationsschätzungen sofort zu aktualisieren, kann eine solche Systemänderung in einem kurzen Zeitraum zahlreiche Ausnahmen bewirken. Diese Ausnahmen würden jedoch im wesentlichen alle aus demselben Grund auftreten, und die Aufsichtsbehörde dürfte dann anders reagieren als bei 10 Ausnahmen, von denen jede eine andere Ursache hat. Die Aufsichtsbehörde könnte dann z.B. einfach verlangen, dass die Bank so rasch wie möglich die Systemänderung in ihrem Modell berücksichtigt, ohne aber die Integrität ihrer Verfahren für die Aktualisierung des Modells zu gefährden.

Auf jeden Fall aber vertritt der Ausschuss den Standpunkt, dass dieser Ausnahmefall nur unter ganz aussergewöhnlichen Umständen zulässig ist; in allen anderen Fällen spricht er sich bei Backtesting-Ergebnissen in der roten Zone ganz entschieden für eine automatische Erhöhung der Eigenkapitalanforderung der Bank aus, ohne einen Ermessensspielraum offenzulassen.

## IV. Schlussbemerkungen

Das beschriebene Rahmenkonzept soll ein einheitliches Vorgehen für die Einbeziehung des Backtesting in die Eigenkapitalberechnung für das Marktrisiko anhand bankeigener Modelle sicherstellen. Dabei wurde das Ziel verfolgt, angemessene und notwendige Anreize in ein Rahmenkonzept einzubauen, das sich sehr stark auf die eigenen Bemühungen der Banken bei der Berechnung ihrer Risiken abstützt; dabei sollen die Grenzen der verfügbaren Instrumente berücksichtigt und der Arbeitsaufwand und die Kosten der vorgeschriebenen Verfahren auf einem Minimum gehalten werden.

Nach Ansicht des Basler Ausschusses für Bankenaufsicht wurde mit dem hier präsentierten Rahmenkonzept diesbezüglich der richtige Mittelweg gefunden. Für vielleicht noch wichtiger hält es der Ausschuss jedoch, dass dieses Konzept den ersten - und damit wesentlichen - Schritt in Richtung einer stärkeren Integration der aufsichtlichen Richtlinien mit nachprüfbarer Messungen des Erfolgs einer Bank darstellt.

Tabelle 1

Genaues Modell			
Anzahl Ausnahmen (von 250 Werten)	Erfassungsquote: 99%		
	exakt	Fehler Typ 1	Fehler Typ 2
0	8,1 %	100,0 %	0,0 %
1	20,5 %	91,9 %	0,6 %
2	25,7 %	71,4 %	3,3 %
3	21,5 %	45,7 %	2,0 %
4	13,4 %	24,2 %	14,0 %
5	6,7 %	10,8 %	17,7 %
6	2,7 %	4,1 %	17,7 %
7	1,0 %	1,4 %	14,8 %
8	0,3 %	0,4 %	10,5 %
9	0,1 %	0,1 %	6,5 %
10	0,0 %	0,0 %	3,6 %
11	0,0 %	0,0 %	1,0 %
12	0,0 %	0,0 %	11
13	0,0 %	0,0 %	12
14	0,0 %	0,0 %	13
15	0,0 %	0,0 %	14

Ungenaues Modell: mögliche andere Erfassungsniveaus			
Anzahl Ausnahmen (von 250 Werten)	Erfassungsquote: 98%	Erfassungsquote: 97%	Erfassungsquote: 96%
	exakt	Fehler Typ 2	exakt
0	0,6 %	0,0 %	0,0 %
1	3,3 %	0,6 %	0,0 %
2	8,3 %	3,9 %	0,4 %
3	14,0 %	12,2 %	3,8 %
4	17,7 %	26,2 %	7,2 %
5	17,7 %	43,9 %	10,9 %
6	14,8 %	61,6 %	13,8 %
7	10,5 %	76,4 %	14,9 %
8	8,0 %	86,9 %	14,0 %
9	6,5 %	86,9 %	14,0 %
10	5,0 %	93,4 %	11,6 %
11	4,0 %	97,0 %	8,6 %
12	3,0 %	98,7 %	5,8 %
13	2,0 %	99,5 %	3,6 %
14	1,0 %	99,8 %	2,0 %
15	0,0 %	99,9 %	1,1 %
	0,0 %	100,0 %	0,5 %

**Annotationen:** Die Tabelle gibt sowohl die exakte Wahrscheinlichkeit an, mit der sich - bei Annahme verschiedener tatsächlicher Erfassungsniveaus - aus einer Stichprobe von 250 unabhängigen Werten eine bestimmte Zahl von Ausnahmen ergibt, als auch die aus diesen exakten Wahrscheinlichkeiten abgeleitete Wahrscheinlichkeit, dass es zu Fehlern des Typs 1 oder 2 kommt.

Der linke Teil der Tabelle gilt für den Fall, dass das Modell genau ist und seine tatsächliche Erfassungsquote 99 % beträgt. Die Wahrscheinlichkeit, dass ein bestimmter Wert eine Ausnahme ist, beträgt somit 1 % (100 % - 99 % = 1 %). Die mit "exakt" bezeichnete Spalte gibt die Wahrscheinlichkeit an, dass sich in einer Stichprobe von 250 Werten genau die bei dieser Annahme genannte Anzahl Ausnahmen ergibt. Die Spalte "Fehler Typ 1" zeigt, mit welcher Wahrscheinlichkeit irrtümlich ein genaues Modell abgelehnt wird, wenn die jeweils angegebene Anzahl Ausnahmen von der Stichprobe von 250 Werten als Schwellen für die Rückweisung eines Modells verwendet wird. Wird die Schwelle z.B. bei 5 oder mehr Ausnahmen angesetzt, besteht gemäß den Angaben in dieser Spalte eine Wahrscheinlichkeit von 10,8 %, dass zu Unrecht ein genaues Modell zurückgewiesen wird.

Im rechten Teil der Tabelle sind vier ungenaue Modelle aufgeführt, deren tatsächliches Erfassungsniveau 98 %, 97 %, 96 % bzw. 95 % beträgt. Bei jedem dieser ungenauen Modelle zeigt die Spalte "exakt", mit welcher Wahrscheinlichkeit sich in einer Stichprobe von 250 Werten genau die bei dieser Annahme genannte Anzahl Ausnahmen ergibt. Die Spalte "Fehler Typ 2" zeigt, mit welcher Wahrscheinlichkeit irrtümlich ein ungenaues Modell mit der angegebenen Erfassungsquote zugelassen wird, wenn die jeweils angegebene Anzahl Ausnahmen von der Stichprobe von 250 Werten als Schwellen für die Rückweisung eines Modells verwendet wird. Wird die Schwelle z.B. bei 5 oder mehr Ausnahmen angesetzt, besteht gemäß den Angaben in dieser Spalte eine Wahrscheinlichkeit von 12,8 %, dass zu Unrecht ein Modell mit einer so niedrigen Erfassungsquote zugelassen wird.

**Tabelle 2**

Zone	Anzahl Ausnahmen	Erhöhung des Multiplikationsfaktors	Kumulative Wahrscheinlichkeit
Grüne Zone	0	0,00	8,11 %
	1	0,00	28,58 %
	2	0,00	54,32 %
	3	0,00	75,81 %
	4	0,00	89,22 %
Gelbe Zone	5	0,40	95,88 %
	6	0,50	98,63 %
	7	0,65	99,60 %
	8	0,75	99,89 %
	9	0,85	99,97 %
Rote Zone	10 und darüber	1,00	99,99 %

**Anmerkungen:** In dieser Tabelle werden die grüne, die gelbe und die rote Zone definiert, anhand deren die Aufsichtsbehörden die Backtesting-Ergebnisse beurteilen, wenn der Eigenkapitalbedarf zur Unterlegung der Marktrisiken mittels interner Modelle berechnet wird. Die Abgrenzungen in dieser Tabelle beruhen auf einer Stichprobe von 250 Werten. Bei anderen Stichprobengrößen beginnt die gelbe Zone an dem Punkt, an dem die kumulative Wahrscheinlichkeit gleich oder höher als 95 % ist, und die rote Zone beginnt an dem Punkt, an dem die kumulative Wahrscheinlichkeit gleich oder höher als 99,99 % ist.

Die kumulative Wahrscheinlichkeit ist die Wahrscheinlichkeit, dass sich in einer Stichprobe von 250 Werten höchstens die angegebene Anzahl Ausnahmen ergibt, wenn das tatsächliche Erfassungsniveau 99 % beträgt. Die z.B. bei 4 Ausnahmen angezeigte kumulative Wahrscheinlichkeit ist die Wahrscheinlichkeit, dass sich zwischen 0 und 4 Ausnahmen ergeben.

Zu beachten ist, dass die Addition dieser kumulativen Wahrscheinlichkeiten und der in Tabelle 1 angezeigten Wahrscheinlichkeiten des Fehlers vom Typ 1 nicht 1 ergibt, denn die kumulative Wahrscheinlichkeit für jede Zahl von Ausnahmen schliesst die Möglichkeit ein, dass sich genau diese Zahl von Ausnahmen ergibt, und das gleiche gilt für die jeweils angegebene Wahrscheinlichkeit des Fehlers vom Typ 1. Die Summe dieser beiden Wahrscheinlichkeiten übersteigt somit 1 um die Höhe der Wahrscheinlichkeit, dass sich genau diese Zahl von Ausnahmen ergibt.