

Beurteilung von Wahrscheinlichkeitsprognosen durch explorative Diagnoseverfahren

Petra Haferkorn und Gerhard Stahl
Bundesaufsichtsamt für das Kreditwesen*,
Berlin

Abstract

Stichworte: Risikomanagement, Value-at-Risk, Diagnostics zur Beurteilung der Güte von Wahrscheinlichkeitsprognosen, Mincer-Zarnowitz, SAS-Insight

1. Einleitung

Seit den Achtziger Jahren entwickelte sich die Modellierung und Analyse von Finanzzeitreihen zu einen wichtigen Zweig innerhalb der Mathematik und Stochastik. Die Gründe hierfür liegen zum einen in der ökonomischen Bedeutung dieser Zeitreihen - man denke etwa an Devisenkurse aber auch Finanzinstrumente wie Optionen auf Aktien - und zum anderen in der Eleganz der zugrunde liegenden mathematischen und stochastischen Modelle und Theorie, Pagan (1996), Bollerslev et al (1992) und Muselia & Rutkowski (1997). Seit Anfang der Neunziger Jahre sehen sich unter anderem Banken einem härteren internationalen Wettbewerb ausgesetzt, der zu einer effizienten Verwendung des zur Verfügung stehenden Eigenkapitals zwingt. Diese Tatsache, aber auch die Notwendigkeit, aus immer komplexeren Instrumenten und Anlagestrategien, Risiken zu identifizieren, zu messen und zu steuern, motiviert die Verwendung stochastischer Modelle in einem portfolioorientierten – also multivariaten – Kontext. Die Publikation von RiskMetrics (1996), dem technischen Dokument des Risikomodells von JP Morgan, New York, gab weltweit wesentliche Impulse zur Anwendung stochastischer Methoden im Risikomanagement. Anlageentscheidungen werden unter Unsicherheit getroffen; somit kommt einer möglichst exakten Quantifizierung des Risikos bei gegebener Information eine Schlüsselrolle zu. Diesem Zweck dienen Risikomodelle. Sie erlauben, monetäre Risiken in Portfolios durch Prognose eines Quantils der Verteilung der Portfoliowertverluste (Value-at-Risk) in aggregierter Form abzuschätzen. Wengleich die unterschiedlichen Modelle, Methoden sowie Fragestellungen wesentlich vom Prognosehorizont abhängen, ist allen stochastischen Risikomodeln eigen, dass sie Wahrscheinlichkeitsprognosen über Umweltzustände (Marktpreise von Instrumenten) mit spezifischen Verlustfunktionen (Risikoprofilen) zu einer Wahrscheinlichkeitsprognose zukünftiger Wertänderungen kombinieren. Die Forderung nach einer einfach zu kommunizierenden, aber dennoch quantitativen Beurteilung der Prognosegüte solcher Modelle liegt auf der Hand, wenn diese Prognosen verlässliche Entscheidungshilfen für das Management sein sollen. Im hier betrachteten Kontext von statistischen Prognosemodellen können Defizite der Wahrscheinlichkeitsprognosen mittels geeigneter Diagnoseverfahren identifiziert werden. Im einfachsten Fall bildet eine univariate Zeitreihe aus Prognosen und Beobachtungen einen wohlbestimmten Gaußschen White-Noise-Prozeß. Daher sind folgende Fragen, die auf Partikularkriterien verweisen, von besonderer Bedeutung:

1. Ist das Modell gut oder schlecht kalibriert?
2. Stammen die Daten aus einer homogenen Population?
3. Sind die Daten unabhängig oder bestehen zeitliche Abhängigkeitsstrukturen?

* Die in diesem Beitrag geäußerten Ansichten und Meinungen sind die persönlichen der Autoren und können nicht als die Meinung des Bundesaufsichtsamtes für das Kreditwesen betrachtet werden.

Im Hinblick auf aktuelle Vorschläge im Rahmen von Kreditrisikomodellen, untersuchen wir die Prognosegüte von Marktrisikomodellen mittels des Kriteriums von Mincer-Zarnowitz:

4. Liefern die hier betrachteten Prognosemodelle rationale Prognosen im Sinne von Mincer- Zarnowitz?

Im Mittelpunkt dieses Beitrages stehen Datenanalysen – vor allem mit SAS/Insight - von Fallbeispielen aus der Praxis. Diese explorativen und interaktiven Techniken wie Box-Plots, Q-Q-Plots, P-P-Plots, Time Plots und Dichteschätzer erlauben zum einen eine Überprüfung obiger Eigenschaften und sind zum anderen geeignet, die Robustheit der getroffenen Schlüsse gegenüber Ausreißern zu untersuchen. Die hier vorgeschlagenen Analyseverfahren lassen nicht nur eine qualitative Aussage darüber zu, ob das Risiko etwa konservativ geschätzt wird, sondern ermöglichen eine detaillierte Quantifizierung der Prognosegüte des Modells.

Nach dieser Einleitung informiert der zweite Abschnitt über den statistischen Rahmen der Value-at-Risk Schätzung. Der dritte Abschnitt stellt die Diagnoseverfahren bereit, die im anschließenden Abschnitt ihre Anwendung auf Fallbeispiele finden. Eine Zusammenfassung beschließt die Arbeit.

2. Statistische Fragestellung

Im Gegensatz zur Optionspreistheorie, bei der stetige Preisprozesse X_t , nämlich die geometrisch Brownsche Bewegung, die Basis bilden, übernehmen bei Risikomodellen Zeitreihen, sogenannte stetige Renditen, Y_t täglicher Preise diese Rolle ($t = 1, \dots, n$):

$$Y_t = \frac{X_{t+1} - X_t}{X_t}.$$

Im einfachsten, univariaten Fall wird ein White-Noise-Prozeß

$$(1) \quad Y_t \stackrel{iid}{\sim} F$$

unterstellt.

Bei VaR-Modellen handelt es sich typischerweise um Prognosemodelle, die jeden Tag eine Vorhersage über die Verteilung

$$(2) \quad \otimes$$

der Wertänderungen des jeweiligen Portfolios treffen. Etwaige weitere Parameter, wie der VaR, werden dann in einem nachgelagerten zweiten Schritt bestimmt. Um eine möglichst sparsame Notation zu verwenden, bezeichnet, wie auch in (1), Y_t fortan einen Vektor stetiger Renditen.

Unter der fundamentalen Annahme, dass die Wertänderungen des Portfolios Π_t eine Funktion T des Renditevektors Y_t ist:

$$(3) \quad v_{t+1}(\Pi_t) - v_t(\Pi_t) = T(Y_{t+1}),$$

wird die (bedingte) Prognoseverteilung \otimes bestimmt.

Beispiel: Der Delta-Normal Ansatz von JP Morgan

Für den einfachen Fall eines Aktienportfolios ist für eine Konkretisierung von (3) einerseits eine Spezifikation eines Verteilungsmodelle für Y_t notwendig und andererseits eine Funktion T anzugeben. In der Praxis hat sich das Modell der multivariaten Normalverteilung bewährt, also

$$Y_t \sim N_d\left(0, \Sigma\right),$$

wobei die Kovarianzmatrix $\hat{\Sigma}_t = \frac{1}{n} X^T X$, mit X der Datenmatrix, geschätzt wird. Bezeichnet

$\lambda_i \in \mathbb{R}$ das Investment in Aktie i , so ergibt sich die Wertänderung durch

$$\lambda_i x_{t+1} - \lambda_i x_t = \lambda_i (x_{t+1} - x_t) = \lambda_i x_t \frac{x_{t+1} - x_t}{x_t} = \lambda_i x_t y_{t+1}.$$

Für $\lambda_i x_t$ wird der Ausdruck „Exposure“, w_i verwendet. Im multivariaten Fall ergibt sich analog

$$v_{t+1}(\Pi_t) - v_t(\Pi_t) = w_t^T Y_{t+1}.$$

T ist in diesem Fall also eine lineare Funktion des Vektors Y_t .

Dann folgt:

$$(4) \quad T_t(Y_t) \sim N(0, \sigma_t^2) \quad \text{mit } \sigma_t^2 \in [0, \infty) \text{ der Portfoliovarianz.}$$

Das Prognoseproblem in (2) kann insofern auf die Schätzung von σ_t^2 , der Portfoliovarianz, reduziert werden.

Als Industriestandard hat sich der VaR als das einseitige 99%-Quantil (manchmal auch 95%-Quantil) der Verteilung in (2) herausgebildet und kann in diesem Fall über

$$(4) \quad VaR_t \approx 2,33 \cdot \hat{\sigma}_t$$

geschätzt werden.

3. Beurteilung der Prognosegüte

In der Praxis ist eine einfache Umsetzung von (2) und (3) – etwa die Schätzung von Σ_t in unserem Beispiel – ohne weitere Vereinfachung nicht möglich. Diese betreffen vor allem die Anzahl der Risikofaktoren im Modell, d.h. die Dimension von Y_t , da sonst $d \gg n$ wäre. Daneben kommen weitere Approximationen zur Anwendung. Hier sind vor allem Taylorapproximation für die Funktion T zu nennen.

Bei der Beurteilung der Prognosegüte von VaR-Modellen sind zwei Aspekte von besonderer Bedeutung. Zum einen ist ein selten gestellte Prognoseproblem, nämlich die Prognose einer ganzen Verteilung im Gegensatz zu dem häufig untersuchten Problem der Punktprognose zu beurteilen. Zum anderen treten klassische Verfahren (z.B. Tests) in den Hintergrund, da lediglich die Adäquanz nicht jedoch die Korrektheit im Sinne eines wahren Modells zur Disposition steht. Aufgrund dieser Vereinfachungen und der damit einhergehenden Fehler wurde das folgende Diagnoseverfahren entwickelt.

4. Beurteilung von Wahrscheinlichkeitsprognosen durch Diagnostics

Obgleich die nachfolgenden Diagnoseverfahren sich lediglich auf das im Rahmen des Beispiels dargelegte Modell beziehen, steht einer Verallgemeinerung auf andere Modellklassen nicht in Frage, wenngleich eine Reihe technischer Hürden zu überwinden sind.

Ausgangspunkt aller Verfahren zur Beurteilung der Prognosegüte bildet die Folge von Paaren aus Prognosen P_t und assoziierten Realisationen I_t

$$(P_t, I_t).$$

Da bei Risikomodellen die Vorhersageregel bekannt ist, liegt es nahe, deren Eigenschaften aus der Verteilung der Prognosegüte zu verwenden. Im folgenden betrachten wir das Beispiel des Delta-Normal Ansatzes, für den

$$(5) \quad S = \frac{L_t}{\text{VaR}} \stackrel{\text{iid}}{\sim} N(0, 2.33^{-2}), \text{ mit } L_t = T(Y_t)$$

gilt. Als Partikularkriterien fungieren:

1. Kalibration
2. Homogenität
3. Unabhängigkeit

die durch folgende Diagnoseverfahren überprüft werden:

1. Liegt für VaR-Modelle eine Folge von Verteilungsfunktionen F_t der Variablen $T(Y_t)$ sowie dazugehörige Portfoliowertänderungen I_t vor, so heißt das Modell gut kalibriert, wenn die Folge $F_t(I_t)$ Realisationen einer Gleichverteilung auf $[0,1]$ sind. Diese Eigenschaft prüfen wir an Hand von P-P-Plots und Q-Q-Plots.
2. Zur Analyse der Homogenitätseigenschaft bedient man sich Q-Q-Plots und Dichteschätzer.
3. Zusätzliches Analysetool sollten Timeplots zur Überprüfung der Resolutions-eigenschaft des Gaußschen White-Noise-Prozesses sein. Eine „gute Resolution“ bezieht sich auf Unabhängigkeitseigenschaften (iid) und liegt vor, wenn keine zeitlichen Strukturen oder Muster in Werten größer (einem Vielfachen) der Standardabweichung erkennbar sind.

5. Rationalität der Prognose nach Mincer-Zarnowitz

Lopez und Saidenberg [1999] schlagen vor, die Beurteilung der Prognosegüte von Kreditrisikomodelle auf ein von Mincer-Zarnowitz (Clements und Hendry, 1998) herrührendes Regressionsverfahren zu gründen. Als Ergänzung zu den bereits dargestellten Diagnoseverfahren wird deren Verfahren hier auch auf Marktrisikomodelle angewendet. Die Grundidee des Verfahrens ist einfach. Sie nutzt die Tatsache, dass im Rahmen des Delta-Normalansatzes das zweite Moment der Variablen L_t gerade die Prognosevarianz ergibt: $E(L_t^2) = \sigma_t^2$.

Identifiziert man den Parameter σ_t^2 mit der Verteilungsfunktion Φ_t , so ist der Bezug zur Prognoseverteilung und den Beobachtungen evident, da I_t^2 als Schätzer für σ_t^2 dient.

Eine Prognose $\hat{\sigma}_t^2$ heißt schwach rational, wenn der Prognosefehler $e_t = I_t^2 - \hat{\sigma}_t^2$ nicht mit der zum Zeitpunkt, wann die Prognose erstellt wird, verfügbaren Informationen prognostizierbar ist.

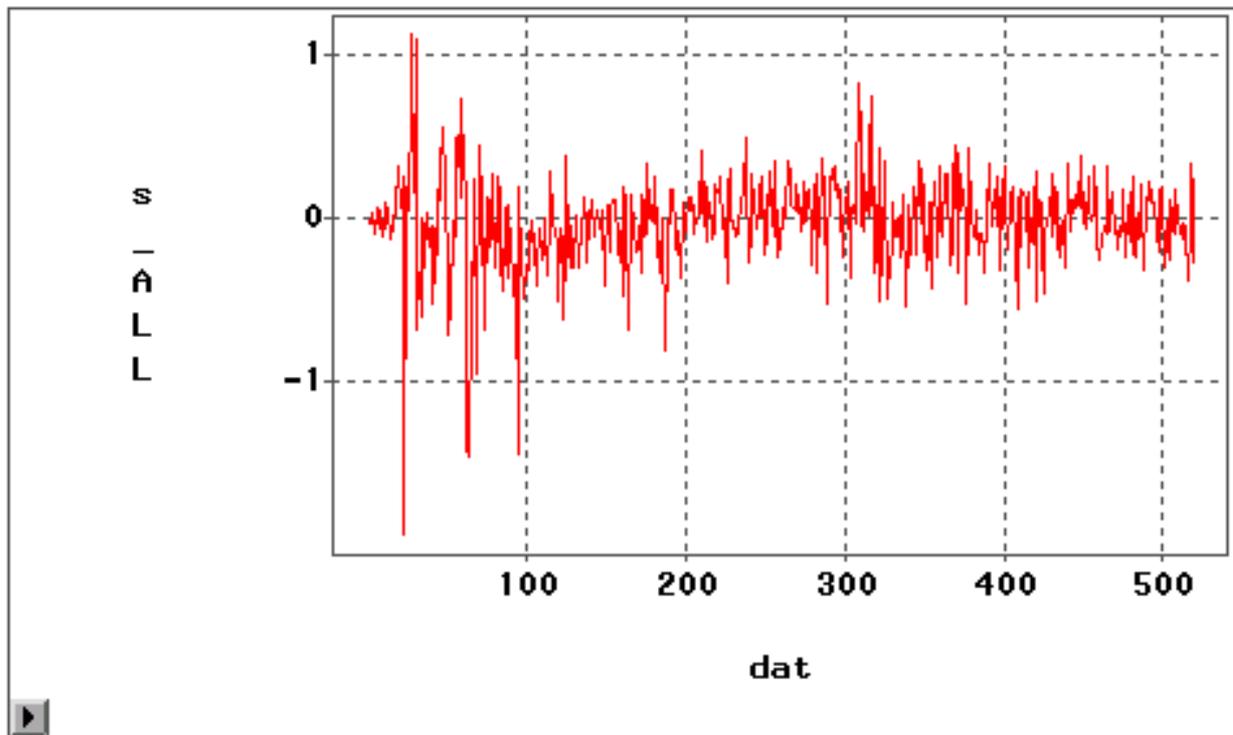
Wir gehen auf einen Teilaspekt ein und untersuchen den Prognosefehler $e_t = I_t^2 - \hat{\sigma}_t^2$ im Hinblick auf systematische Fehler. Dreidimensionale Timeplots ergänzen diese Untersuchungen.

6. Fallstudien

Es liegen die Zeitreihen zweier Banken A und B vor, für A der letzten beiden Jahre vom Umfang 522 Werte und für B vom letzten Jahr mit dem Umfang 201 Werte. Gegeben sind die hypothetischen Portfoliowertänderungen und die Risikoschätzung auf Gesamtbankebene und für wichtige Teilportfolien. Aufgrund der Normalverteilung der Gewinne und Verluste erübrigt sich eine Edgeworthentwicklung, obgleich zur Modellierung des VaR jeweils eine Simulation verwendet wurde.

Beurteilung von Wahrscheinlichkeitsprognosen durch Diagnostics

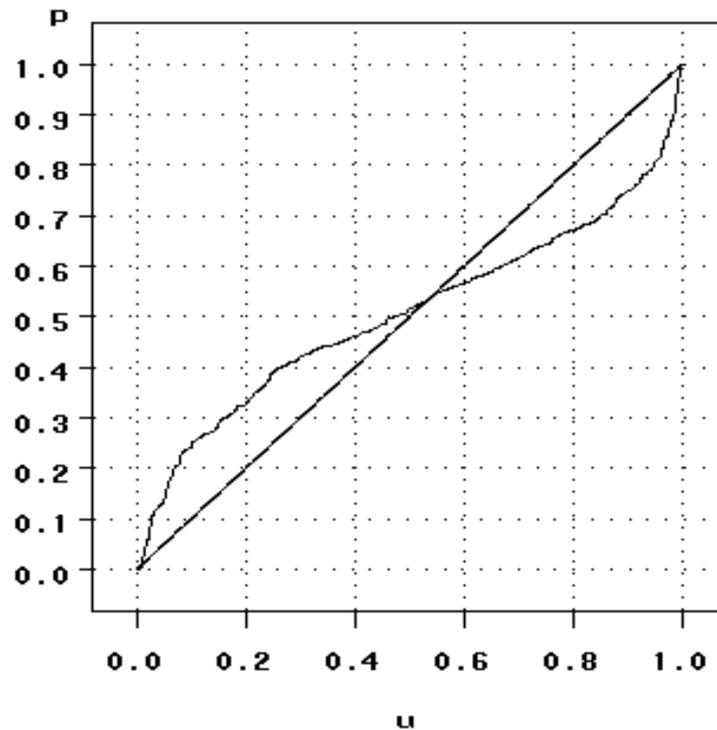
Angesichts der Notwendigkeit einer Anonymisierung der Daten, wird auf einen Timeplot der Portfoliowertänderungen und der Risikoschätzung verzichtet.



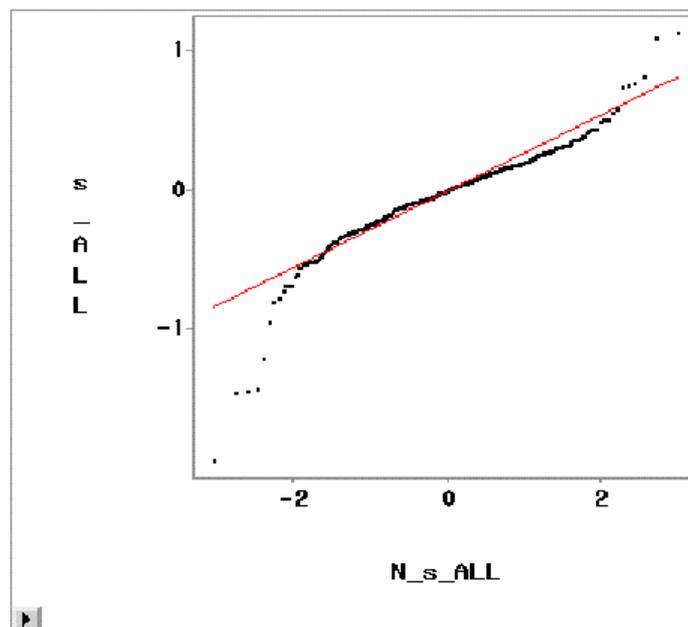
Figur 1: Timeplot des Quotienten s für das Gesamtbankportfolio

Auffällig sind in Figur 1 die anfänglichen Überschreitungen (Quotient s größer Eins) und die Konzentration der sich anschließenden Werte (insgesamt 493 von 522 Beobachtungen) im Intervall -0.5 und 0.5 . Die Überschreitungen sind nicht durch Portfolioumschichtungen, sondern einen sich erweiternden Spread zwischen zwei Zinskurven bedingt, also implizit einer Unterdrückung eines Risikofaktors. Die nachfolgende geringe Amplitude der Zeitreihe lässt eine deutliche Überschätzung des Risikos aufgrund einer schlechten Kalibration des VaR-Modells vermuten. Bereits an dieser Graphik lässt sich die Heterogenität ablesen.

P-P-Plot



Figur 2: P-P-Plot der empirischen Percentile des Quotienten s gegen die Gleichverteilung
 Der P-P-Plot weist auf die Symmetrieeigenschaft der Verteilung hin und unterstreicht die zu konservative Risikoschätzung (Figur 2).



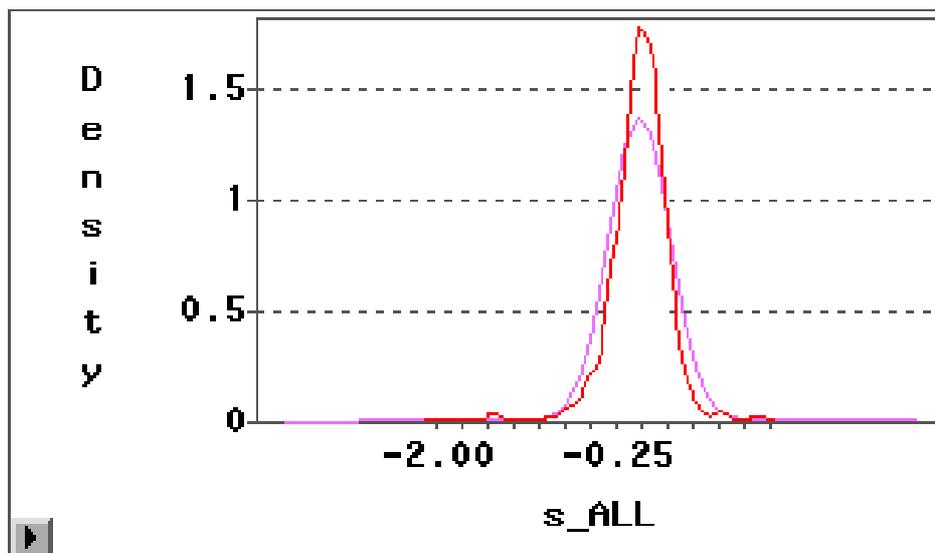
Figur 3: Q-Q-Plot der empirischen Quantile des Quotienten s gegen die Standardnormalverteilung

Die in Figur 3 dargelegten Heterogenitäten sind für die Praxis eher typisch als eine Ausnahme. Dies gilt ebenfalls für die damit einhergehenden Ausreißer. Bekanntlich haben letztere einen großen Einfluss bei der Schätzung von σ durch den üblichen Momentenschätzer, was in der Folge die Beurteilung der Konservativität des Modells ebenfalls stark beeinflusst.

Moments			
N	522.0000	Sum Wgts	522.0000
Mean	-0.0031	Sum	-1.6030
Std Dev	0.2890	Variance	0.0835
Skewness	-1.2423	Kurtosis	7.8374
USS	43.5183	CSS	43.5134
CV	-9410.7674	Std Mean	0.0126

Tabelle 1: Momente der geschätzten Verteilung von s

Die empirische Standardabweichung beträgt 0.289. Ein Vergleich mit der Standardabweichung aus dem Prognosemodell ($2,33^{-1} \approx 0,429$) ergibt eine durchschnittliche Überschätzung um fast 50% (Tabelle 1).



Figur 4: Dichteschätzung von s mit parametrischer und Kernschätzung

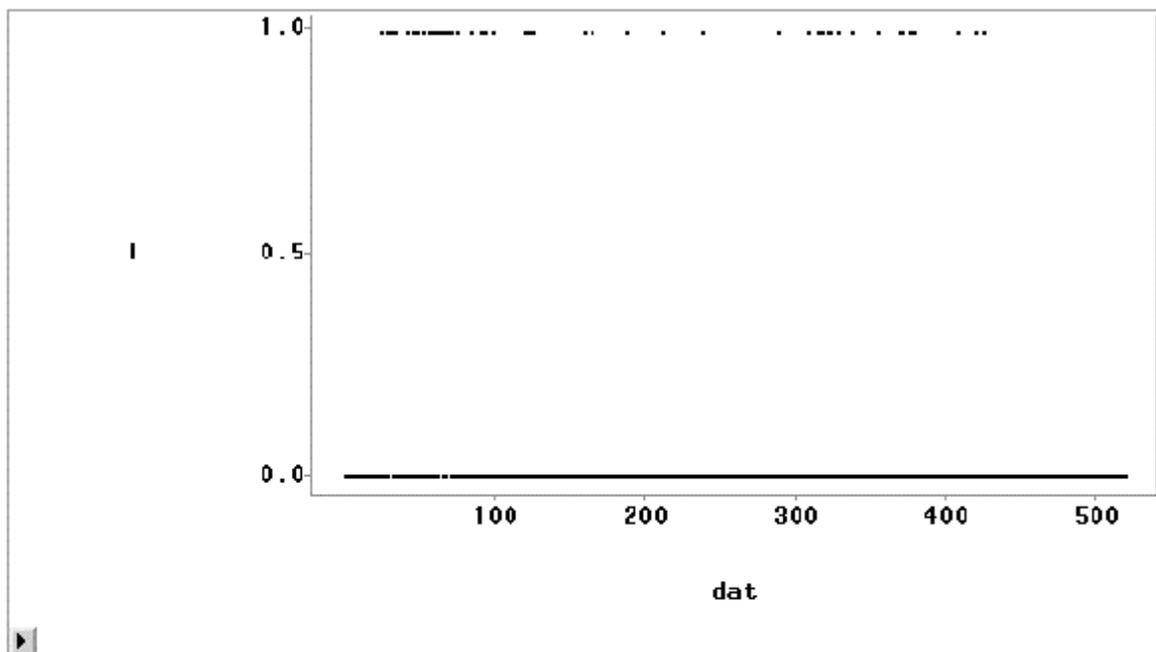
Daher werden die Effekte der Ausreißer durch eine graphische Veranschaulichung einer geschätzten Dichte (Figur 4) einer Normalverteilung mit der aus einer nichtparametrischen Dichteschätzung sichtbar gemacht. Zusätzlich dienen robuste Schätzungen der Varianz einer zahlenmäßigen Quantifizierung.

Robust Measures of Scale		
Measure	Value	Estimate of Sigma
Interquartile Range	0.2769	0.2053
Gini's Mean Difference	0.2932	0.2598
MAD	0.1391	0.2063
Sn	0.2195	0.2195
Qn	0.2290	0.2273

Tabelle 2: Robuste Schätzungen der Standardabweichung von s

Die getroffenen Feststellungen können in Tabelle 2 anhand von Zahlen für robuste Schätzungen der Standardabweichung noch mal nachvollzogen werden.

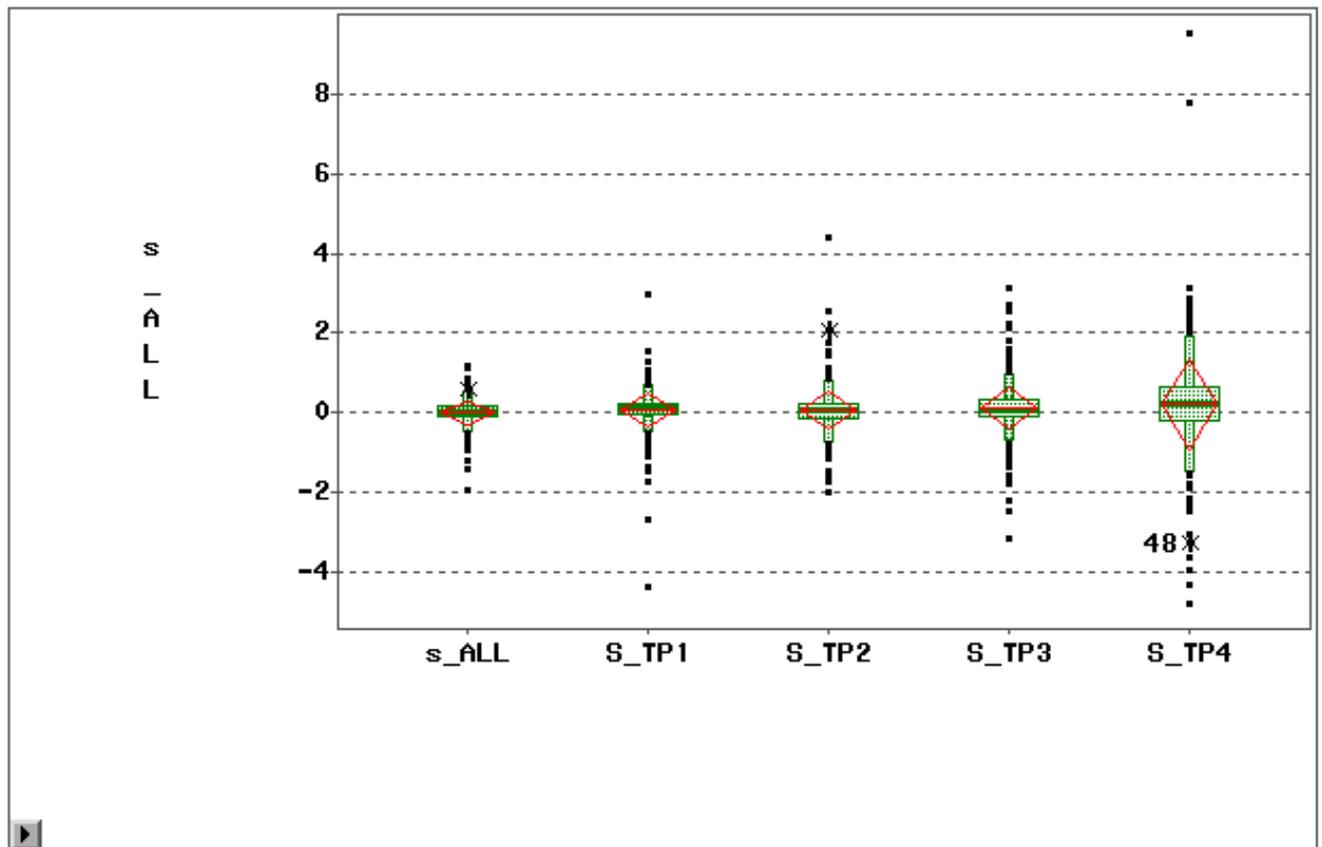
Eine detaillierte Untersuchung der homogenen Perioden der Zeitreihe ergab, dass für die erste keine wesentliche Unterschätzung des Risikos, wohingegen im zweiten Teil eine Risikoüberschätzung um 100% auf einer Datenbasis von 450 Beobachtungen vorliegt. Darüber hinaus ergaben sich keine besonderen Auffälligkeiten.



Figur 5: Timeplot der Realisationen von $|S| > 0.4$

Der Plot der absoluten Überschreitungen des Schwellenwertes 0.4 lässt – insbesondere zu Beginn der Zeitreihe, aber auch im Bereich zwischen 300 und 400 - exponierte zeitliche Abhängigkeitsstrukturen erkennen (Figur 5). Die Ursachen hierfür sind vielschichtig. In der Praxis kommen neben inkorrekten Inputdaten vor allem fehlende Risikofaktoren oder eine nicht adäquate Approximation in Frage.

Die Analyse der Teilportfolios dient der Identifikation von Ursachen der Risikoüberschätzung auf Gesamtbankebene bzw. der Entdeckung von Masking-effekten, d.h. Fehler unterschiedlicher Vorzeichen heben sich nach der Aggregation auf der nächsten Stufe auf. Boxplots dienen als effizientes Hilfsmittel zur Durchführung der simultanen Analyse von Gesamtbank- und Teilportfolios (Figur 6).



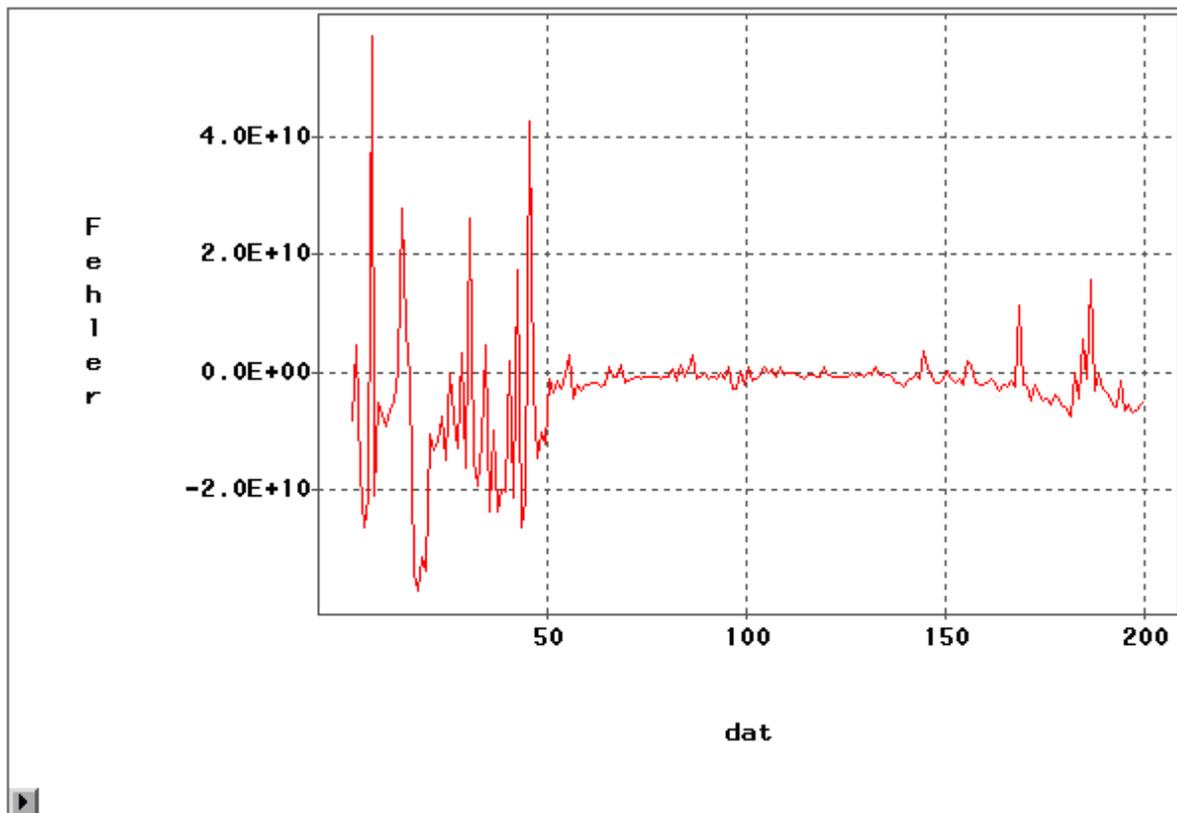
Figur 6: Boxplots von s für ein Portfolio und vier Teilportfolios

Die mit einem Kreuz markierten Beobachtungen – aber auch die insgesamt sehr extremen Beobachtungen - in Figur 6 zeigen solche Masking-effekte. Die Analyse der Unter- bzw. Überschätzung auf Teilportfolioebene ergab, dass das Teilportfolio 4 das Risiko deutlich unterschätzte.

Rationalität der Prognose nach Mincer-Zarnowitz

In diesem Abschnitt untersuchen wir den zweiten Datensatz der Bank B, der bereits an anderer Stelle untersucht wurde. Im Zuge dieser Analysen wurde das Prognosemodell als sehr gut eingestuft; das Risiko wurde durch das Modell beinahe 100%ig genau geschätzt.

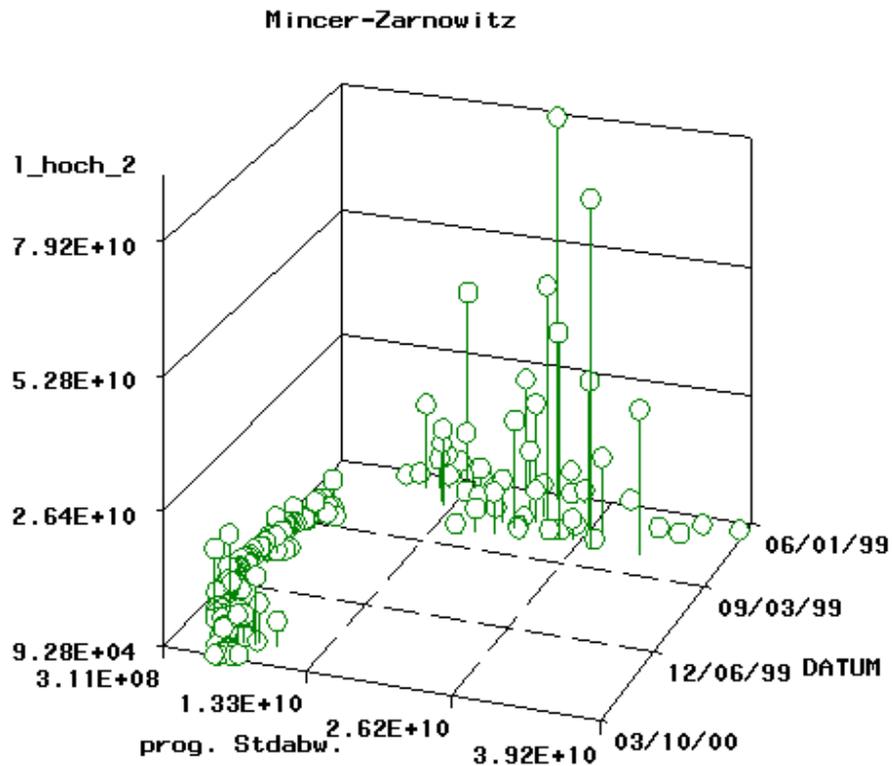
Eine erste Analyse gilt dem Fehlerprozess e_{t+1} .



Figur 7: Der Fehlerprozess e_{t+1} des neuen Datensatzes

Der Timeplot von e_{t+1} in Figur 7 ist annähernd symmetrisch zur x-Achse. Insoweit stützen die Daten, dass der Prognosefehler im Mittel den Wert Null annimmt.

Die folgende Graphik (Figur 8) enthält den Plot des Vektor $(\hat{\sigma}_t^2, 1_t^2)_{t=1}^{201}$. Trotz einer Reihe unterschiedlicher Rotationen, war es nicht möglich, eine lineare Beziehung visuell herzustellen. Entsprechende Berechnungen von Parametern (z.B. R^2) ergaben nicht das klare Bild welches der Prognosegüte des Modells entsprochen hätte. Insoweit enttäuschte die Anwendung der Mincer-Zarnowitz Regression.



Figur 8: Dreidimensionale Darstellung der Werte 1_{t+1}^2 (Achse „1 hoch 2“) und $\hat{\sigma}_t^2$ (Achse „prog. Varianz“) in Abhängigkeit der Zeit (Achse „datum“)

Zusammenfassung

Die hier vorgestellten Verfahren zur Beurteilung der Prognosegüte von VaR-Modellen haben sich in der Praxis sowohl von Banken als auch bei der Aufsicht bewährt. Diese einfachen Methoden erlauben es auch Nicht-Statistikern - nach einer kurzen Einarbeitungszeit - eine sichere Beurteilung der Prognosegüte. Darüber hinaus gewähren detaillierte Analysen über die gesamte Hierarchie des Portfolios genaue Einblicke in die Ursachen für Unter- oder Überschätzungen.

Literatur

- Bollerslev, T.; Chou, R. Y. and K. F. Kroner (1992) ARCH Modeling in Finance *J. Econometrics* **52** : 2-59
- Clements, M. und Hendry, D. (1998): Forecasting Economic Time Series, Cambridge University Press
- Lopez/Saidenberg (1999) Evaluating Credit Risk Models. Research Paper, Federal Reserve Bank, New York
- Muselia, M. and M. Rutkowski (1997) Martingale Methods in Financial Modelling
- Overbeck, L. und Stahl, G. (2000) Backtesting: Allgemeine Theorie, Praxis und Perspektiven, Handbuch Risikomanagement, Band 1, Seiten 289-321, Herausgeber L. Johanning und B. Rudolph, Uhlenbruch Verlag Bad Soden im Taunus
- Pagan, A. (1996) The Econometrics of Financial Markets *J. Emp. Finance* **3** : 15-102
- RiskMetrics (1996) Technical Dokument, 4th Ed.