

# **Dichteprognose autoregressiver Zeitreihen**

Jürgen Jacobs

Fakultät III - Umwelt und Technik, Leuphana Universität Lüneburg

## **Inhaltsverzeichnis**

<b>1</b>	<b>Einleitung und Themenabgrenzung .....</b>	<b>2</b>
<b>2</b>	<b>Univariate autoregressive Zeitreihenmodelle .....</b>	<b>3</b>
<b>3</b>	<b>Vektor-autoregressive Zeitreihenmodelle .....</b>	<b>7</b>
<b>4</b>	<b>Prognose .....</b>	<b>13</b>
<b>5</b>	<b>Anwendungsbeispiel .....</b>	<b>16</b>
	<b>Literatur .....</b>	<b>22</b>

## **1 Einleitung und Themenabgrenzung**

Zeitreihenprognosen versuchen die künftige Entwicklung von interessierenden Größen wie bspw. Zinsen, Inflationsraten oder Aktienkursen aus Kenntnis historischer Werte zu extrapolieren. Im univariaten Fall wird nur eine Größe betrachtet, im multivariaten Fall werden Interdependenzen zwischen mehreren Größen berücksichtigt. Während im univariaten Fall künftige Werte alleine aus den historischen Werten extrapoliert werden, könnten im multivariaten Fall auch kausale Modelle wie z. B. Zinstheorien Berücksichtigung finden. Wir beschränken uns hier auf reine Zeitreihenmodelle, d. h., die Interdependenzen zwischen den betrachteten Größen werden alleine mit Hilfe historischer Werte ermittelt. Ebenso sei angemerkt, dass Zeitreihen nicht der einzig denkbare Gegenstand einer Prognose sind. Ebenso könnte man Prognosen von Ereignissen betrachten, wie das Eintreten eines Ereignisses mit ungewissem Ausgang (z. B. Ausgang einer Wahl) oder den Zeitpunkt des Eintretens eines sicheren Ereignisses (z. B. Zinsänderung).

In der Zeitreihenprognose lassen sich drei Prognosetypen unterscheiden: Punkt-, Intervall- und Dichteprognose. Bei der Punktprognose wird nur ein einzelner Wert (in der Regel der bedingte Erwartungswert bei gegebenen historischen Werten) geschätzt. Bei der Intervallprognose wird ein Wertebereich geschätzt, in welchen der Prognosewert mit einer bestimmten Wahrscheinlichkeit liegen wird. Bei der Dichteprognose wird die gesamte Wahrscheinlichkeitsverteilung des Prognosewerts geschätzt. In der vorliegenden Arbeit werden Dichteprognosen auf der Basis von Simulationsrechnungen betrachtet.

## 2 Univariate autoregressive Zeitreihenmodelle

Eine wichtige Klasse der univariaten Zeitreihenmodelle bilden die sog. autoregressiven Modelle. Sie beschreiben die aktuelle Beobachtung als gewichtete Summe zurückliegender Realisationen. Ein autoregressiver Prozess der Ordnung  $p$  (AR( $p$ )-Prozess) folgt der Beziehung:

$$(1) \quad x_t = \alpha + \varphi_1 \cdot x_{t-1} + \varphi_2 \cdot x_{t-2} + \dots + \varphi_p \cdot x_{t-p} + u_t, \quad t = p, p+1, p+2, \dots$$

Die Störgrößen  $u_t$  werden als unabhängige Zufallsvariable angenommen, die alle dieselbe Verteilung (independent and identically distributed (IID)) um den Mittelwert 0 haben:  $u_t \sim IID(0, \sigma^2)$ . Gleichung (1) beschreibt eine Regression der Variablen  $x_t$  auf ihre eigenen verzögerten Werte. Dies erklärt die Namensgebung.

Eine wichtige Voraussetzung für die Parameterschätzung ist die Stationarität der Zeitreihe. Eine Zeitreihe heißt *streng stationär*, falls die gemeinsame Verteilung von  $(x_{t+l_1}, x_{t+l_2}, \dots, x_{t+l_k})$  für beliebig gewählte  $l_1, l_2, \dots, l_k$  nur von den Verschiebungen  $l_1, l_2, \dots, l_k$  und nicht vom Zeitpunkt  $t$  abhängt. Dies ist in der Praxis selten gegeben. Daher wird in der Zeitreihenanalyse meist nur eine schwache Stationarität angenommen. Eine Zeitreihe heißt *schwach stationär*, falls gilt:

$$(a) \quad E(x_t) = \mu,$$

$$(b) \quad C(x_t, x_{t-l}) = \gamma_l.$$

Das heißt, der Erwartungswert ( $E$ ) ist zeitunabhängig, und die Autokovarianz ( $C$ ) hängt nur von der Verschiebung  $l$ , nicht aber vom Zeitpunkt  $t$  ab.

Ein AR(1)-Modell ist gegeben durch:

$$(2) \quad x_t = \alpha + \varphi_1 \cdot x_{t-1} + u_t, \quad t = 1, 2, 3, \dots$$

Nehmen wir die Gültigkeit von (2) für beliebig weit zurück liegende Zeitpunkte  $t = -1, -2, -3, \dots$  an, so ergibt die sukzessive Anwendung von Gleichung (2) für  $x_{t-1}, x_{t-2}, \dots$ :

$$(3) \quad x_t = \alpha + u_t + \varphi_1(\alpha + u_{t-1}) + \varphi_1^2(\alpha + u_{t-2}) + \dots$$

Falls  $|\varphi_1| < 1$ , beschreibt (3) einen schwach stationären Prozess. Für den Erwartungswert erhält man:

$$\mu = E(x_t) = \alpha + \varphi_1\alpha + \varphi_1^2\alpha + \dots = \alpha \sum_{i=0}^{\infty} \varphi_1^i = \frac{\alpha}{1 - \varphi_1}.$$

Dabei wurde  $E(u_t) = 0$  verwendet. Die Beziehung  $|\varphi_1| < 1$  wurde für die Konvergenz der Reihe benötigt. Für die Autokovarianz erhält man wegen  $E(u_i \cdot u_i) = \sigma^2$  und  $E(u_i \cdot u_j) = 0$  für  $i \neq j$ :

$$\begin{aligned} C(x_t, x_{t-l}) &= E[(x_t - \mu)(x_{t-l} - \mu)] \\ &= E[(u_t + \varphi_1 u_{t-1} + \varphi_1^2 u_{t-2} + \dots + \varphi_1^l u_{t-l} + \varphi_1^{l+1} u_{t-l-1} + \varphi_1^{l+2} u_{t-l-2} + \dots)(u_{t-l} + \varphi_1 u_{t-l-1} + \varphi_1^2 u_{t-l-2} + \dots)] \\ &= (\varphi_1^l + \varphi_1^{l+2} + \varphi_1^{l+4} + \dots)\sigma^2 = \varphi_1^l (1 + \varphi_1^2 + \varphi_1^4 + \dots)\sigma^2 = \frac{\varphi_1^l}{1 - \varphi_1^2} \sigma^2. \end{aligned}$$

Im Falle von  $\varphi_1 = 1$  spricht man von einem sog. *Random-Walk-Prozess*. Die sukzessive Anwendung von (2) auf  $x_{t-1}, x_{t-2}, \dots, x_1$  liefert:

$$(4) \quad x_t = x_0 + \alpha \cdot t + \sum_{i=1}^t u_i.$$

Erwartungswert und Autokovarianz sind nun zeitabhängig:

$$E(x_t) = x_0 + \alpha \cdot t, \quad C(x_t, x_{t-l}) = (t-l) \cdot \sigma^2.$$

Der Teil  $x_0 + \alpha \cdot t$  beschreibt ein deterministisches Wachstumsverhalten. Man spricht von daher von einem *stochastischen Trend*. Charakteristisch für Random-Walk-Prozesse mit oder ohne Trend ( $\alpha = 0$ ) ist ein aufgrund der kumulierten Störgrößen unbeschränktes Wachstum mit der Zeit:  $Var(x_t) = t \cdot \sigma^2$ .

Die kumulierten Störgrößen können zu längeren Phasen wachsender und fallender Werte von  $x_t$  führen, die ein trendartiges Verhalten vortäuschen und daher bei Regressionsanalysen zu fehlerhaften Modellen führen. Bereits Granger und Newbold (1974) haben auf dieses sog. *Spurious-Regression-Problem* hingewiesen. Um Random-Walk-Prozesse im Rahmen der klassischen Zeitreihenanalyse behandeln zu können, werden Differenzen gebildet. Ist beispielsweise eine Zeitreihe  $x_t$  ein Random-Walk-Prozess ohne Trend, so ist die erste Differenz  $\Delta x_t = x_t - x_{t-1} = u_t$  eine Zeitreihe, die durch die Störgröße beschrieben wird. Nun ist denkbar, dass die ersten Differenzen ebenfalls ein Random-Walk-Verhalten aufweisen. In diesem Fall werden zweite Differenzen gebildet:  $\Delta^2 x_t = \Delta(x_t - x_{t-1}) = x_t - 2 \cdot x_{t-1} + x_{t-2}$ . Dieses Verfahren wird so lange fortgesetzt, bis man kein Random-Walk-Verhalten mehr antrifft. Ein Prozess, bei dem  $d$  Differenzen zu bilden sind, heißt *integriert* von der Ordnung  $d$  oder kurz  $I(d)$ -Prozess.

Betrachten wir einen AR(1)-Prozess ohne Interzept  $\alpha$ :  $x_t = \varphi_1 x_{t-1} + u_t$ . Die gewöhnliche Kleinste-Quadrate-Schätzung (ordinary least squares (OLS)-Schätzung) liefert:

$$\hat{\varphi}_1 = \frac{\sum_{t=1}^n x_t \cdot x_{t-1}}{\sum_{t=1}^n x_{t-1}^2}$$

Zum Testen der Nullhypothese  $\varphi_1 = 1$  bietet sich die Statistik

$$\tau = \frac{\hat{\varphi}_1 - 1}{std(\hat{\varphi}_1)}$$

an;  $std(\hat{\varphi}_1)$  bezeichnet den Standardfehler des Schätzers  $\hat{\varphi}_1$ . Die Verteilung von  $\tau$  hängt vom wahren Wert  $\varphi_1$  ab. Für  $|\varphi_1| < 1$  folgt  $\tau$  der t-Verteilung mit n Freiheitsgraden. Im Falle eines Random Walks – also bei Zutreffen der Nullhypothese – folgt  $\tau$  nicht der t-Verteilung. Kritische Schranken können mit Hilfe von Monte-Carlo-Simulationen ermittelt werden. Sie liegen unterhalb der Schranken der t-Verteilung mit unendlich vielen Freiheitsgraden. Dies zeigt, dass die Nullhypothese bei Entscheidung auf Basis einer t-Verteilung viel zu oft verworfen würde. Der Test auf Basis der  $\tau$ -Statistik wird *Unit-root-Test* oder nach den Entdeckern *Dickey-Fuller-Test* genannt.

Das Modell  $x_t = \varphi_1 x_{t-1} + u_t$  lässt sich umschreiben zu  $\Delta x_t = \delta x_{t-1} + u_t$  mit  $\delta = \varphi_1 - 1$ . Ein Random Walk ist demnach nicht auszuschließen, falls die Hypothese  $\delta = 0$  nicht abgelehnt werden kann. Im Fall eines AR(p)-Prozesses mit  $p > 1$  lässt sich das von Dickey und Fuller (1979) vorgeschlagene Testverfahren in seiner verallgemeinerten Form verwenden. Zusätzlich können noch ein Interzept und ein deterministischer Trend in die Modellgleichung aufgenommen werden:

$$(5) \quad \Delta x_t = \delta x_{t-1} + \beta_1 \Delta x_{t-1} + \dots + \beta_{p-1} \Delta x_{t-p-1} + u_t,$$

$$(6) \quad \Delta x_t = \alpha + \delta x_{t-1} + \beta_1 \Delta x_{t-1} + \dots + \beta_{p-1} \Delta x_{t-p-1} + u_t,$$

$$(7) \quad \Delta x_t = \alpha + \beta \cdot t + \delta x_{t-1} + \beta_1 \Delta x_{t-1} + \dots + \beta_{p-1} \Delta x_{t-p-1} + u_t.$$

Gleichung (6) geht aus Gleichung (1) hervor, indem man  $\delta = \varphi_1 + \varphi_2 + \dots + \varphi_p - 1$  und

$\beta_i = -\sum_{j=i+1}^p \varphi_j$  setzt. Die Dickey-Fuller t-Statistiken zur Überprüfung der Nullhypothese

$\delta = \varphi_1 + \varphi_2 + \dots + \varphi_p - 1 = 0$ , dass ein Random-Walk für die Modelle (5), (6) bzw. (7)

anzunehmen ist, werden mit  $\tau$ ,  $\tau_\mu$  bzw.  $\tau_\tau$  bezeichnet. Mit Hilfe des F-Tests nach Dickey und Fuller (1981) kann die Signifikanz der Parameter  $\alpha$  bzw.  $\beta$  überprüft werden:

$$F = \frac{S_0 - S}{S} \cdot \frac{n+1-l}{r}.$$

$S_0$  bezeichnet die Summe der quadrierten Residuen bei Zutreffen der Nullhypothese,  $S$  die Summe der quadrierten Residuen ohne Restriktion,  $r$  die Anzahl der Parameterrestriktionen bei Zutreffen der Nullhypothese,  $n + 1$  den Stichprobenumfang und  $l$  die Anzahl der Parameter. Die entsprechenden Teststatistiken nennt man  $\Phi_1$  (Modell (6), Nullhypothese:  $\alpha = \delta = 0$ ),  $\Phi_2$  (Modell (7), Nullhypothese:  $\alpha = \beta = \delta = 0$ ) bzw.  $\Phi_3$  (Modell (7), Nullhypothese:  $\beta = \delta = 0$ ).

Die Anzahl  $p$  der zu berücksichtigenden Verzögerungen wird üblicherweise mit Hilfe von sog. Informationskriterien ermittelt, die den Schätzfehler mit der Anzahl  $l$  der verwendeten Parameter zu einer Größe verknüpfen. Je mehr Parameter verwendet werden, desto „leichter“ ist es, den Schätzfehler klein zu halten. Daher wird jede Hinzunahme von Parametern „bestraft“. Häufig verwendet werden die *Informationskriterien* von Akaike (1973) (*AIC*) und Schwarz (1978) (*SIC*). Bis auf konstante Terme sind die Kriterien wie folgt definiert:

$$AIC = \log\left(\frac{1}{n-p+1} \sum_{i=p}^n \hat{u}_i^2\right) + \frac{2}{n-p+1} l,$$

$$SIC = \log\left(\frac{1}{n-p+1} \sum_{i=p}^n \hat{u}_i^2\right) + \frac{\log(n-p+1)}{n-p+1} l,$$

wobei  $\hat{u}_i$  die durch die Residuen geschätzten Störgrößen bezeichnen. Die Ordnung  $p$  wird nun so bestimmt, dass das gewählte Kriterium einen minimalen Wert annimmt. Für  $n - p + 1 > 7$  gilt  $\log(n - p + 1) > 2$ ; daher führt die Verwendung des Schwarzschen Informationskriteriums im Allgemeinen zu einer niedrigeren Ordnung als die Verwendung des Informationskriteriums von Akaike.

### 3 Vektor-autoregressive Zeitreihenmodelle

Zur Berücksichtigung von kontemporären und intertemporären Abhängigkeiten zwischen den untersuchten Zeitreihen bieten sich Vektor-autoregressive (VAR) Modelle an. Bei  $k$  Zeitreihen und der Berücksichtigung von  $p$  Verzögerungen erhält man folgendes Gleichungssystem:

$$(8) \quad \begin{aligned} y_{1,t} &= \alpha_1 + \delta_{11,1} \cdot y_{1,t-1} + \dots + \delta_{1k,1} \cdot y_{k,t-1} + \dots + \delta_{11,p} \cdot y_{1,t-p} + \dots + \delta_{1k,p} \cdot y_{k,t-p} + u_{1,t}, \\ y_{2,t} &= \alpha_2 + \delta_{21,1} \cdot y_{1,t-1} + \dots + \delta_{2k,1} \cdot y_{k,t-1} + \dots + \delta_{21,p} \cdot y_{1,t-p} + \dots + \delta_{2k,p} \cdot y_{k,t-p} + u_{2,t}, \\ &\dots \\ y_{k,t} &= \alpha_k + \delta_{k1,1} \cdot y_{1,t-1} + \dots + \delta_{kk,1} \cdot y_{k,t-1} + \dots + \delta_{k1,p} \cdot y_{1,t-p} + \dots + \delta_{kk,p} \cdot y_{k,t-p} + u_{k,t}. \end{aligned}$$

Falls einzelne Zeitreihen ein Random-Walk-Verhalten aufweisen, ist es naheliegend, diese zunächst durch Differenzenbildung in stationären Zeitreihen zu überführen und anschließend in das Gleichungssystem (8) aufzunehmen. Dadurch werden jedoch möglicherweise zu viele Differenzen gebildet – nämlich dann, wenn Kointegrationsbeziehungen zwischen einzelnen Zeitreihen bestehen. *Kointegration* bedeutet, dass eine Linearkombination nicht-stationärer Zeitreihen gefunden werden kann, die stationär ist. Das Konzept der Kointegration geht auf Granger (1981) und Engle und Granger (1987) zurück. Tiao, Tsay und Wang (1993) zeigen an einem Beispiel auf, dass es problematisch sein kann, existierende Gleichgewichtsbeziehungen mit einem Kointegrationstest zu erkennen. Insbesondere wird von ihnen kritisiert, dass die Größe der Streuung der einzelnen Zeitreihen vom Kointegrationstest nicht berücksichtigt wird. Die Bedeutung von Kointegrationsbeziehungen für langfristige Prognosen wurde u. a. von Engle und Yoo (1987) herausgestellt. Christoffersen und Diebold (1997) bemängeln allerdings zu Recht, dass die Ergebnisse schwer zu interpretieren sind, da nicht das VAR-Modell der stationären Zeitreihen mit dem Kointegrationsmodell verglichen wird, sondern das VAR-Modell der nicht-stationären Niveauwerte. In einer eigenen Untersuchung kommen sie zu dem Resultat, dass die Berücksichtigung von Kointegrationsbeziehungen für langfristige Prognosen zu schlechteren Ergebnissen führt als die Verwendung univariater integrierter Variablen. Im Folgenden soll daher auf eine Betrachtung von Kointegrationsmodellen verzichtet werden. Die Bedeutung von Unit-Root-Tests und Kointegrationstests für die Spezifikation von VAR-Modellen ist in der Literatur umstritten. Einen aktuelle Studie zur Modellauswahl und einen Literaturüberblick findet man bei Allen und Fildes (2007).

Ein weiteres Problem von VAR-Modellen ist, dass die Anzahl der zu schätzenden Parameter rasch mit der Dimension  $k$  und der Ordnung  $p$  anwächst. Beispielsweise sind bei 5 Zeitreihen und 3 Verzögerungen bereits  $k \cdot (1 + k \cdot p) = 5 \cdot (1 + 5 \cdot 3) = 80$  Parameter zu schätzen. Eine hohe Anzahl von Parametern erschwert nicht nur die Interpretation des Modells, sondern führt auch zu großen Ungenauigkeiten bei der Parameterschätzung, falls die Anzahl der verfügbaren historischen Zeitreihenwerte nicht erheblich größer als die Anzahl der Parameter ist. In der Praxis ist häufig eine große Anzahl der geschätzten Parameter nicht signifikant von Null verschieden. Zur Elimination nicht signifikanter Parameter bietet sich das sequentielle Verfahren nach Brüggemann und Lütkepohl (2001) an, welches nacheinander die Regressoren eliminiert, die zur größten Reduktion des Informationskriterium führen, bis keine weitere Reduktion mehr möglich ist.

Für eine Stichprobe mit Indexmenge  $I = \{0, 1, 2, \dots, n\}$  ergibt sich in Matrixform folgendes Gleichungssystem:

$$(9) \quad Y = \delta \cdot Z + U,$$

wobei  $\delta$  die vollständige  $k \times (1 + k \cdot p)$ -Parametermatrix

$$\delta = \begin{pmatrix} \alpha_1 & \delta_{11,1} & \cdots & \delta_{1k,1} & \cdots & \delta_{11,p} & \cdots & \delta_{1k,p} \\ \alpha_2 & \delta_{21,1} & \cdots & \delta_{2k,1} & \cdots & \delta_{21,p} & \cdots & \delta_{2k,p} \\ & & & \ddots & & & & \\ \alpha_k & \delta_{k1,1} & \cdots & \delta_{kk,1} & \cdots & \delta_{k1,p} & \cdots & \delta_{kk,p} \end{pmatrix}$$

bezeichnet und

$$Y = [Y_p \cdots Y_n], Y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix}, Z = [Z_{p-1} \cdots Z_{n-1}], Z_t = \begin{bmatrix} 1 \\ Y_t \\ Y_{t-1} \\ \cdots \\ Y_{t-p+1} \end{bmatrix}, U = [U_p \cdots U_n], U_t = \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \cdots \\ u_{k,t} \end{bmatrix}.$$

Die Restriktionen lassen sich mit Hilfe einer geeigneten Matrix  $R$  in der Form  $\text{vec}(\delta) = R\gamma$  notieren, wobei  $\gamma$  den Vektor der von Null verschiedenen Parameter bezeichnet. Der  $\text{vec}$ -Operator überführt eine Matrix in einen Spaltenvektor, welcher – beginnend mit der ersten Spalte – aus den untereinander geschriebenen Spalten der Matrix besteht. Sei zum Beispiel  $k = 2$  und  $p = 2$ , so wird aus dem Gleichungssystem (8):

$$\begin{aligned} y_{1,t} &= \alpha_1 + \delta_{11,1} \cdot y_{1,t-1} + \delta_{12,1} \cdot y_{2,t-1} + \delta_{11,2} \cdot y_{1,t-2} + \delta_{12,2} \cdot y_{2,t-2} + u_{1,t}, \\ y_{2,t} &= \alpha_2 + \delta_{21,1} \cdot y_{1,t-1} + \delta_{22,1} \cdot y_{2,t-1} + \delta_{21,2} \cdot y_{1,t-2} + \delta_{22,2} \cdot y_{2,t-2} + u_{2,t}. \end{aligned}$$

Die Restriktion  $\delta_{21,1} = 0$  lässt sich dann in der Form

$$\text{vec}(\delta) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \delta_{11,1} \\ 0 \\ \delta_{12,1} \\ \vdots \\ \delta_{22,2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ & & & \ddots & & \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \delta_{11,1} \\ \delta_{12,1} \\ \vdots \\ \delta_{22,2} \end{pmatrix}$$

schreiben. Für die nachfolgenden Umformungen werden folgende Rechenregeln für passende Matrizen verwendet:

- (R1)  $(A \otimes B)' = A' \otimes B'$ .
- (R2)  $(A \otimes B)(C \otimes D) = AC \otimes BD$ .
- (R3)  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ .
- (R4)  $\text{vec}(AB) = (B' \otimes I)\text{vec}(A)$ .

Hierbei bezeichnen  $I$  die Einheitsmatrix,  $\otimes$  das Kronecker-Produkt und  $A'$  die transponierte Matrix  $A$ . Das Kronecker-Produkt einer  $(m \times n)$ -Matrix  $A = (a_{ij})$  und  $(p \times q)$ -Matrix  $B = (b_{ij})$  ergibt eine  $(mp \times nq)$ -Matrix, die wie folgt definiert ist:

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ & \ddots & \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}.$$

Aus (9) und  $\text{vec}(\delta) = R\gamma$  folgt:

$$(10) \quad \text{vec}(Y) = \text{vec}(\delta \cdot Z) + \text{vec}(U) \stackrel{(R4)}{=} (Z' \otimes I_k)\text{vec}(\delta) + \text{vec}(U) = (Z' \otimes I_k)R\gamma + \text{vec}(U).$$

Für das obige Beispiel mit  $k = 2, p = 2$  und der Restriktion  $\delta_{21,1} = 0$  ergibt sich:

$$\begin{aligned}
\text{vec}(Y) &= \begin{pmatrix} y_{1,2} \\ y_{2,2} \\ \vdots \\ y_{1,n} \\ y_{2,n} \end{pmatrix} = \begin{bmatrix} 1 & y_{1,1} & y_{2,1} & y_{1,0} & y_{2,0} \\ 1 & y_{1,2} & y_{2,2} & y_{1,1} & y_{2,1} \\ & & \ddots & & \\ 1 & y_{1,n-1} & y_{2,n-1} & y_{1,n-2} & y_{2,n-2} \end{bmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \delta_{11,1} \\ 0 \\ \delta_{12,1} \\ \vdots \\ \delta_{22,2} \end{pmatrix} + \begin{pmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{1,n} \\ u_{2,n} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & y_{1,1} & 0 & \cdots & y_{2,0} & 0 \\ 0 & 1 & 0 & y_{1,1} & \cdots & 0 & y_{2,0} \\ & & \ddots & & & & \\ 1 & 0 & y_{1,n-1} & 0 & \cdots & y_{2,n-2} & 0 \\ 0 & 1 & 0 & y_{1,n-1} & \cdots & 0 & y_{2,n-2} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \delta_{11,1} \\ 0 \\ \delta_{12,1} \\ \vdots \\ \delta_{22,2} \end{pmatrix} + \begin{pmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{1,n} \\ u_{2,n} \end{pmatrix}.
\end{aligned}$$

Zur Parameterschätzung sollte eine verallgemeinerte Kleinste-Quadrate-Anpassung (generalized least squares (GLS)-Schätzung) durchgeführt werden, da diese bei restringierten Modellen effizienter ist als eine OLS-Schätzung ist (vgl. Lütkepohl (2005, S. 71, 199)). Für die Kovarianzmatrix von  $\text{vec}(U)$  gilt:

$$C_{\text{vec}(U)} = I_{n-p+1} \otimes C_u,$$

wobei  $C_u$  die zeitlich konstante Kovarianzmatrix der Störgrößen  $u_t$  bezeichnet. Nach dem GLS-Verfahren ist

$$S(\gamma) = \text{vec}(U)' C_{\text{vec}(U)}^{-1} \text{vec}(U) = [\text{vec}(Y) - (Z' \otimes I_k) R \gamma]' (I_{n-p+1} \otimes C_u)^{-1} [\text{vec}(Y) - (Z' \otimes I_k) R \gamma]$$

zu minimieren. Mit obigen Rechenregeln ergeben sich folgende Vereinfachungen beim Ausmultiplizieren von  $S(\gamma)$ :

$$(I_{n-p+1} \otimes C_u)^{-1} \stackrel{(R3)}{=} I_{n-p+1} \otimes C_u^{-1},$$

$$[(Z' \otimes I_k) R \gamma]' \stackrel{(R1)}{=} [R \gamma]' (Z' \otimes I_k)' = \gamma' R' (Z \otimes I_k),$$

$$(I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) \stackrel{(R2)}{=} I_{n-p+1} Z' \otimes C_u^{-1} I_k = Z' \otimes C_u^{-1},$$

$$(Z \otimes I_k) (I_{n-p+1} \otimes C_u^{-1}) \stackrel{(R2)}{=} Z I_{n-p+1} \otimes I_k C_u^{-1} = Z \otimes C_u^{-1},$$

$$(Z \otimes I_k) (I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) \stackrel{(R2)}{=} (Z \otimes C_u^{-1}) \otimes (Z' \otimes I_k) = Z Z' \otimes C_u^{-1}.$$

Man erhält:

$$\begin{aligned}
 S(\gamma) &= [\text{vec}(Y) - (Z' \otimes I_k) R \gamma]' (I_{n-p+1} \otimes C_u^{-1}) [\text{vec}(Y) - (Z' \otimes I_k) R \gamma] \\
 &= \text{vec}(Y)' (I_{n-p+1} \otimes C_u^{-1}) \text{vec}(Y) - \text{vec}(Y)' (I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) R \gamma \\
 &\quad - [(Z' \otimes I_k) R \gamma]' (I_{n-p+1} \otimes C_u^{-1}) \text{vec}(Y) + [(Z' \otimes I_k) R \gamma]' (I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) R \gamma \\
 &= \text{vec}(Y)' (I_{n-p+1} \otimes C_u^{-1}) \text{vec}(Y) - \text{vec}(Y)' (I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) R \gamma \\
 &\quad - \gamma' R' (Z \otimes I_k) (I_{n-p+1} \otimes C_u^{-1}) \text{vec}(Y) + \gamma' R' (Z \otimes I_k) (I_{n-p+1} \otimes C_u^{-1}) (Z' \otimes I_k) R \gamma \\
 &= \text{vec}(Y)' (I_{n-p+1} \otimes C_u^{-1}) \text{vec}(Y) - \text{vec}(Y)' (Z' \otimes C_u^{-1}) R \gamma \\
 &\quad - \gamma' R' (Z \otimes C_u^{-1}) \text{vec}(Y) + \gamma' R' (Z Z' \otimes C_u^{-1}) R \gamma.
 \end{aligned}$$

Ein Minimum liegt vor, wenn  $\frac{\partial S}{\partial \gamma'} = (0, \dots, 0)$  und die Hessische Matrix  $\frac{\partial^2 S}{\partial \gamma \partial \gamma'}$  positiv definit

ist. Unter Verwendung der Produktregel  $\frac{\partial [a(\gamma)' A b(\gamma)]}{\partial \gamma'} = b' A' \frac{\partial a}{\partial \gamma'} + a' A \frac{\partial b}{\partial \gamma'}$  erhalt man die

Gleichung:

$$\frac{\partial S}{\partial \gamma'} = -\text{vec}(Y)' (Z' \otimes C_u^{-1}) R - [R' (Z \otimes C_u^{-1}) \text{vec}(Y)]' + \gamma' [R' (Z Z' \otimes C_u^{-1}) R]' + \gamma' [R' (Z Z' \otimes C_u^{-1}) R].$$

Da  $C_u^{-1}$  symmetrisch ist und  $(Z Z')' = Z Z'$ , gilt wegen Regel (R1):

$$\begin{aligned}
 [R' (Z Z' \otimes C_u^{-1}) R]' &= [R' (Z Z' \otimes C_u^{-1})' R]' = [R' (Z Z' \otimes C_u^{-1}) R], \\
 [R' (Z \otimes C_u^{-1}) \text{vec}(Y)]' &= \text{vec}(Y)' (Z' \otimes C_u^{-1}) R.
 \end{aligned}$$

Daraus folgt:

$$\frac{\partial S}{\partial \gamma'} = -2 \text{vec}(Y)' (Z' \otimes C_u^{-1}) R + 2 \gamma' [R' (Z Z' \otimes C_u^{-1}) R]'.$$

Damit gilt  $\frac{\partial S}{\partial \gamma'} = (0, \dots, 0)$ , falls

$$\begin{aligned}
 \gamma' [R' (Z Z' \otimes C_u^{-1}) R]' &= \text{vec}(Y)' (Z' \otimes C_u^{-1}) R, \\
 \gamma' &= \text{vec}(Y)' (Z' \otimes C_u^{-1}) R ([R' (Z Z' \otimes C_u^{-1}) R]')^{-1}, \\
 \gamma &= [R' (Z Z' \otimes C_u^{-1}) R]^{-1} R' (Z \otimes C_u^{-1}) \text{vec}(Y).
 \end{aligned}$$

Bei der letzten Gleichung wurde die Beziehung  $(A^{-1})' = (A')^{-1}$  verwendet.

Da  $[R' (Z Z' \otimes C_u^{-1}) R]$  symmetrisch ist (s. o.: der Ausdruck ist gleich dem transponierten

Ausdruck), folgt  $\frac{\partial^2 S}{\partial \gamma \partial \gamma'} = 2 \cdot [R' (Z Z' \otimes C_u^{-1}) R]$ . Diese Matrix ist positiv definit, falls die Zeilen

von  $Z$  linear unabhangig sind:

Zunachst ist  $C_u^{-1}$  positiv definit, da  $C_u$  positiv definit ist.  $Z Z' = Z I_{n-p+1} Z'$  ist positiv definit, da

$I_{n-p+1}$  positiv definit ist und ein voller Spaltenrang von  $Z'$  vorausgesetzt wird. Damit ist auch

$ZZ' \otimes C_u^{-1}$  positiv definit. Schließlich ist  $R'(ZZ' \otimes C_u^{-1})R$  positiv definit, da  $R$  den vollen Spaltenrang hat.

Damit ist der GLS-Schätzer für  $\gamma$  durch

$$\hat{\gamma} = [R'(ZZ' \otimes \hat{C}_u^{-1})R]^{-1} R'(Z \otimes \hat{C}_u^{-1}) \text{vec}(Y)$$

gegeben. Als Schätzer für die unbekannte Kovarianzmatrix lässt sich

$$\hat{C}_u = \frac{1}{n - p + 1 - (k \cdot p + 1)} (Y - \hat{\delta}Z)(Y - \hat{\delta}Z)'$$

mit dem OLS-Schätzer  $\hat{\delta} = YZ'(ZZ')^{-1}$  des nicht-restringierten Gleichungssystems (8) verwenden. Bei dem Schätzer der Kovarianzmatrix zum berechneten restringierten System ist dagegen auf die Subtraktion von  $k \cdot p + 1$  zu verzichten, da bei jeder der  $k$  Gleichungen deutlich weniger als  $k \cdot p + 1$  Parameter verwendet werden.

Wie im univariaten Fall kann die Anzahl der Verzögerungen mit Hilfe eines Informationskriteriums ermittelt werden. Bis auf konstante Terme sind die Kriterien wie folgt definiert:

$$AIC = \log \left( \det \left( \frac{1}{n - p + 1} \sum_{i=p}^n \hat{U}_i \hat{U}_i' \right) \right) + \frac{2}{n - p + 1} l,$$

$$SIC = \log \left( \det \left( \frac{1}{n - p + 1} \sum_{i=p}^n \hat{U}_i \hat{U}_i' \right) \right) + \frac{\log(n - p + 1)}{n - p + 1} l,$$

wobei  $\hat{U}_i$  die  $(k \times 1)$ -Spaltenvektoren der Residuen und  $l$  die Anzahl der geschätzten Parameter bezeichnen. Lütkepohl (2005, S. 153 ff.) konnte für ein bivariates Modell die Vorteilhaftigkeit des Schwarzischen Informationskriteriums gegenüber anderen Informationskriterien mit Hilfe einer Simulationsrechnung nachweisen.

#### 4 Prognose

Mit Hilfe der geschätzten Parameter können künftige Werte durch sukzessive Anwendung der Regressionsgleichung ermittelt werden. Im Falle von Gleichung (1) erhält man:

$$(11) \quad \begin{aligned} \hat{x}_{n+1} &= \hat{\alpha} + \hat{\phi}_1 \cdot x_n + \hat{\phi}_2 \cdot x_{n-1} + \dots + \hat{\phi}_p \cdot x_{n-p} + \hat{u}_{n+1}, \\ \hat{x}_{n+2} &= \hat{\alpha} + \hat{\phi}_1 \cdot \hat{x}_{n+1} + \hat{\phi}_2 \cdot x_n + \dots + \hat{\phi}_p \cdot x_{n+1-p} + \hat{u}_{n+2}, \\ &\dots \\ \hat{x}_{n+h} &= \hat{\alpha} + \hat{\phi}_1 \cdot \hat{x}_{n+h-1} + \hat{\phi}_2 \cdot \hat{x}_{n+h-2} + \dots + \hat{\phi}_p \cdot \hat{x}_{n+h-p} + \hat{u}_{n+h}, \text{ für } h > p. \end{aligned}$$

Die Wahrscheinlichkeitsverteilung der Prognosewerte lässt sich über eine Simulation der stochastischen Störgrößen  $u_i$  ermitteln. Dabei werden die statistischen Eigenschaften der unbeobachteten Störgrößen üblicherweise mit Hilfe der Residuen geschätzt. Bei Finanzzeitreihen weisen die ermittelten Residuen häufig eine im Vergleich zur Normalverteilung größere Kurtosis (Wölbung) auf. Dies bedeutet ein häufigeres Auftreten von kleinen Residuen in der Mitte und großen Residuen an den Rändern der Verteilung. Krämer und Runde (2000) zeigen, dass bei der Normalisierung von leptokurtischen Verteilungen eine Anpassung mit empirischen Interquantilbereichen wesentlich robuster ist als eine Anpassung mit der empirischen Standardabweichung. Um diese Effekte zu berücksichtigen, lässt sich mit Hilfe eines Normalverteilungstests prüfen, ob die einzelnen Residuen durch eine an das Maximum aus der empirischen Standardabweichung und dem empirischen Interdezilbereich (Streubreite der mittleren 80% der Fälle) angepasste Normalverteilung mit Mittelwert 0 modelliert werden können:

$$\hat{u}_t \sim \hat{\sigma} \cdot N(0,1); \hat{\sigma} = \sqrt{\frac{n-p+1}{n-p+1-l}} \cdot \text{Max}\left(\hat{s}, \frac{I_{80}}{2,564}\right).$$

Der Quotient 2,564 ist der Interdezilbereich der Standard-Normalverteilung  $N(0,1)$ .  $I_{80}$  bezeichnet den empirisch ermittelten Interdezilbereich,  $n+1$  den Stichprobenumfang,  $l$  die Anzahl der Parameter und  $\hat{s}$  die empirische Standardabweichung.

Für nicht normalverteilte Residuen lässt sich ein symmetrischer *Kerndichteschätzer* mit Gauß-Kern der Form

$$\hat{f}_s(x) = \frac{1}{2}(\hat{f}(x) + \hat{f}(-x)), \text{ mit } \hat{f}(x) = \frac{1}{(n-p+1) \cdot h} \sum_{i=p}^n K\left(\frac{x - \hat{u}_i}{h}\right), K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

mit der Bandbreite wird  $h = 1,06 \cdot \hat{s} \cdot (n-p+1)^{-1/5}$  (vgl. Silverman (2003), S. 45) verwenden, falls mit einem Chi-Quadrat-Test die Nullhypothese (Übereinstimmung des Kerndichteschätzers mit der empirischen Verteilung) nicht abgelehnt werden kann. Alternativ

kann ein *Bootstrap*-Verfahren verwendet werden, bei dem die Verteilungsfunktion der Störgröße durch zufällige Ziehung der Residuen (mit Zurücklegen) simuliert wird. In diesem Fall wird allerdings die Größe der Störung durch den maximalen Wert der Residuen beschränkt.

Neben der korrekten Ermittlung der Verteilungsfunktion der Störgrößen ist für die Prognose wesentlich, dass keine Autokorrelationen vorliegen. Ansonsten führt die sukzessive Anwendung der Regressionsgleichung gemäß (11) zu fehlerhaften Prognosewerten. Autokorrelationen  $\rho_i = \text{Corr}(u_t, u_{t-i})$  bis zu einer bestimmten Ordnung  $m$  können mit der *Q-Statistik* von Ljung und Box (1978)

$$Q_m = (n - p + 1)(n - p + 3) \sum_{i=1}^m \frac{\hat{\rho}_i^2}{n - p + 1 - i} \text{ mit } \hat{\rho}_i = \text{Corr}(\hat{u}_t, \hat{u}_{t-i}).$$

überprüft werden. Die Nullhypothese besagt, dass bis zu einer vorgegebenen Verzögerung keine Autokorrelationen vorliegen:  $\rho_1 = \rho_2 = \dots = \rho_m = 0$ .

Zum Test der Stabilität der Zeitreihenmodelle bietet sich der *Prognosetest nach Chow* (1960) an. Dazu schätzt man ein neues Regressionsmodell für eine bis zu einem vorgegebenen Zeitpunkt verkürzte Zeitreihe mit den selben Regressoren wie das zu prüfende Modell. Die Nullhypothese besagt, dass die Parameter beider Modelle gleich sind. Die Teststatistik hat die Form:

$$F = \frac{S - S_1}{S_1} \cdot \frac{n_1 - l}{n + 1 - n_1},$$

wobei  $n_1$  die Länge der verkürzten Zeitreihe,  $S$  die Summe der quadrierten Residuen,  $S_1$  die Summe der quadrierten Residuen des an die verkürzte Zeitreihe angepassten Modells und  $l$  die Anzahl der Regressoren bezeichnen.

Bei den multivariaten Modellen sind die Prognosewerte analog zu (11) zu berechnen. Da in die Berechnung der Prognosewerte die Störgrößen vorheriger Perioden eingehen, dürfen wie im univariaten Fall keine Autokorrelationen und zusätzlich keine Kreuzkorrelationen mit verzögerten Störgrößen auftreten. Die Verallgemeinerung der Q-Statistik von Ljung und Box für VAR-Modelle ist für die Ordnung  $m$  durch

$$Q_m = (n + 1)^2 \sum_{i=1}^m \frac{1}{n + 1 - i} \text{tr}(\hat{C}_i' \hat{C}_0^{-1} \hat{C}_i' \hat{C}_0^{-1}) \text{ mit } \hat{C}_i = \frac{1}{n + 1} \sum_{j=i+1}^{n+1} \hat{U}_j \hat{U}_{j-i}'$$

gegeben (Lütkepohl, 2004, S. 127). Allerdings sind kontemporäre Abhängigkeitsstrukturen der Störgrößen zu berücksichtigen. Im Falle einer multivariaten Normalverteilung können die Störgrößen über eine *Cholesky-Faktorisierung* der Kovarianzmatrix  $\hat{C}_u = GG'$  berechnet werden:  $\hat{U}_t = G\varepsilon_t$ . Dabei bezeichnen  $G$  eine untere Dreiecksmatrix und  $\varepsilon_t$  einen  $(k \times 1)$ -Vektor mit  $k$  standard-normalverteilten Zufallsvariablen. Eine solche Faktorisierung existiert für alle positiv definiten Matrizen, also auch für die Kovarianzmatrix. Man überzeugt sich leicht, dass  $\hat{U}_t = G\varepsilon_t$  die gewünschte Kovarianzmatrix besitzt:

$$E(\hat{U}_t \hat{U}_t') = E(G\varepsilon_t (G\varepsilon_t)') = E(G\varepsilon_t \varepsilon_t' G') = GE(\varepsilon_t \varepsilon_t')G' = GI_k G' = GG' = \hat{C}_u.$$

Falls keine multivariate Normalverteilung vorliegt, ist die Verwendung der Kovarianzmatrix zur Beschreibung der Abhängigkeitsstrukturen nicht geeignet. Stattdessen bietet sich die *Copula-Methode* an (Embrechts et al. 2005). Alternativ kann wie im univariaten Fall das Bootstrap-Verfahren verwendet werden, wobei bei jeder Ziehung der Residuen derselbe Zeitpunkt für die verschiedenen Zeitreihen zu wählen ist, damit die Abhängigkeitsstrukturen erhalten bleiben.

Eine Verallgemeinerung des Chow-Prognosetests für VAR-Modelle findet man in Lütkepohl (2004, S. 136).

## 5 Anwendungsbeispiel

Nachfolgend sollen univariate und multivariate autoregressive Modelle für die in Tabelle 1 beschriebenen Zeitreihen ermittelt werden.

**Tabelle 1** Zeitreihen

Zeitreihe	Bedeutung	Quelle	Zeitraum
cpir_de	Änderungsrate pro Quartal des Verbraucherpreisindex in Deutschland	Eurostat bis November 2002, Statistisches Bundesamt ab Dezember 2002	Q1 1984 – Q4 2004
cpir_uk	Änderungsrate pro Quartal des Verbraucherpreisindex in Großbritannien (vor 1997 durch das Office for National Statistics geschätzt)	Office for National Statistics	Q1 1984 – Q4 2004

In Tabelle 2 sind die p-Werte der Dickey-Fuller-Tests aufgeführt.

**Tabelle 2** p-Werte der Dickey-Fuller-Tests

Zeitreihe $x$	$\tau_\tau(x)$	$\Phi_3(x)$	$\tau_\mu(x)$	$\Phi_1(x)$	$\tau(x)$	$\tau(\Delta x)$
cpir_de	0,61	> 0,1	0,29	> 0,1	0,29	0,000
cpir_uk	0,49	> 0,1	0,54	> 0,1	0,27	0,000

Die nicht signifikanten p-Werte bezüglich  $\tau_\tau(x)$  und  $\Phi_3(x)$  besagen, dass die Nullhypothese eines Random Walks mit  $\beta = 0$  für Modell (7) nicht zu verwerfen ist. Wegen der nicht signifikanten p-Werte bezüglich  $\tau_\mu(x)$  und  $\Phi_1(x)$  kann die Nullhypothese eines Random Walks mit  $\alpha = 0$  für Modell (6) nicht abgelehnt werden. Daher bleibt Modell (5) zu prüfen. Wegen der nicht signifikanten p-Werte bezüglich  $\tau(x)$  und der hoch-signifikanten p-Werte bezüglich  $\tau(\Delta x)$  können alle Zeitreihen als integriert von der Ordnung 1 ohne linearen Trend  $\beta$  und ohne Interzept  $\alpha$  modelliert werden. Tabelle 3 zeigt die Ergebnisse einer OLS-Schätzung.

**Tabelle 3** AR-Modelle (Standardabweichungen der Parameter in Klammern)

$\Delta\text{cpir\_de}_t = -0,729 (0,100)$ $\Delta\text{cpir\_de}_{t-1} - 0,835 (0,087)$ $\Delta\text{cpir\_de}_{t-2} - 0,472 (0,099)$ $\Delta\text{cpir\_de}_{t-3}$
$\Delta\text{cpir\_uk}_t = -0,925 (0,082)$ $\Delta\text{cpir\_uk}_{t-1} - 0,839 (0,092)$ $\Delta\text{cpir\_uk}_{t-2} - 0,687 (0,080)$ $\Delta\text{cpir\_uk}_{t-3}$

Tabelle 4 fasst die Ergebnisse der Residuenanalyse zusammen. Der Ljung-Box-Test wurde für bis zu 16 Verzögerungen durchgeführt. Der Chow-Prognosetest wurde für alle Zeitpunkte ab dem 1. Quartal 1995 durchgeführt. Für beide Tests sind die minimalen p-Werte angegeben. Wie man sieht, sind auf einem Signifikanzniveau von 5 % keine Autokorrelationen für die Residuen anzunehmen, und die Modelle können nach dem Chow-Prognosetest als stabil betrachtet werden. Zur Simulation der Störtgrößen können nach dem Anderson-Darling-Test (Stephens, 1974) Normalverteilungen verwendet werden.

**Tabelle 4** p-Werte der Residuen-Tests

	Anderson-Darling-Test	Ljung-Box-Test	Chow-Prognosetest	$\hat{\sigma}$
cpir_de	0,77	> 0,44	0,32	0,0041
cpir_uk	0,21	> 0,23	0,54	0,0060

Tabelle 5 zeigt die Ergebnisse der GLS-Schätzungen für das VAR-Modell.

**Tabelle 5** VAR-Modell (Standardabweichungen der Parameter in Klammern)

$\Delta\text{cpir\_de}_t = -0,776 (0,087)$ $\Delta\text{cpir\_de}_{t-1} - 0,801 (0,080)$ $\Delta\text{cpir\_de}_{t-2} - 0,137 (0,045)$ $\Delta\text{cpir\_uk}_{t-2} - 0,428 (0,090)$ $\Delta\text{cpir\_de}_{t-3}$
$\Delta\text{cpir\_uk}_t = -0,908 (0,076)$ $\Delta\text{cpir\_uk}_{t-1} - 0,830 (0,088)$ $\Delta\text{cpir\_uk}_{t-2} - 0,687 (0,075)$ $\Delta\text{cpir\_uk}_{t-3}$

Tabelle 6 fasst die Ergebnisse der Residuenanalyse zusammen. Wie im univariaten Fall wurde der Ljung-Box-Test für bis zu 16 Verzögerungen und der Chow-Prognosetest wurde für alle Zeitpunkte ab dem 1. Quartal 1995 durchgeführt. Für beide Tests sind die minimalen p-Werte angegeben. Wie sieht, muss die Hypothese einer multivariaten Normalverteilung der Residuen nach dem Test von Doornik and Hansen (1994) verworfen werden. Nach dem multivariaten Ljung-Box-Test sind keine Auto- und Kreuzkorrelationen der Residuen für bis zu 16 Verzögerungen anzunehmen, und nach dem Chow-Prognosetest kann das VAR-Modell als stabil betrachtet werden.

**Tabelle 6** p-Werte der Residuen-Tests

	Doornik-Hansen-Test	Ljung-Box-Test	Chow-Prognosetest
(cpir_de, cpir_uk)	0	> 0,34	0,35

Nach 10.000 Simulationsläufen ergeben sich für die univariaten Modelle folgende Werte (Werte in eckigen Klammern beziehen sich auf die mit dem Bootstrap-Verfahren simulierten Störgrößen, die übrigen Werte wurden mit normalverteilten Störgrößen ermittelt):

**Tabelle 7** cpir\_de

	nach einem Quartal	nach 16 Quartalen
$\mu$	0,66 % [0,65 %]	0,58 % [0,60 %]
$\sigma$	0,41 % [0,39 %]	0,68 % [0,65 %]

**Tabelle 8** cpir\_uk

	nach einem Quartal	nach 16 Quartalen
$\mu$	0,11 % [0,10 %]	0,51 % [0,41 %]
$\sigma$	0,60 % [0,56 %]	0,94 % [0,90 %]

Die Abbildungen 1 und 2 verdeutlichen die „Verbreiterung“ der Häufigkeitsverteilung (10.000 Simulationsläufe mit normalverteilten Störgrößen) mit zunehmenden Prognosehorizont.

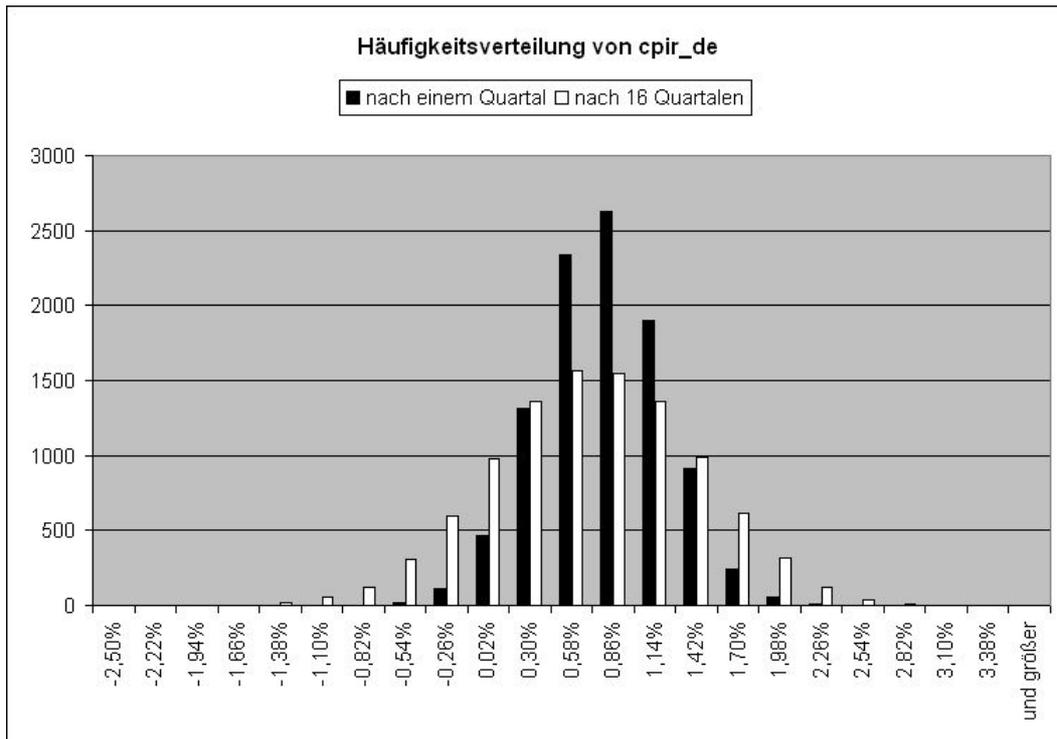


Abbildung 1

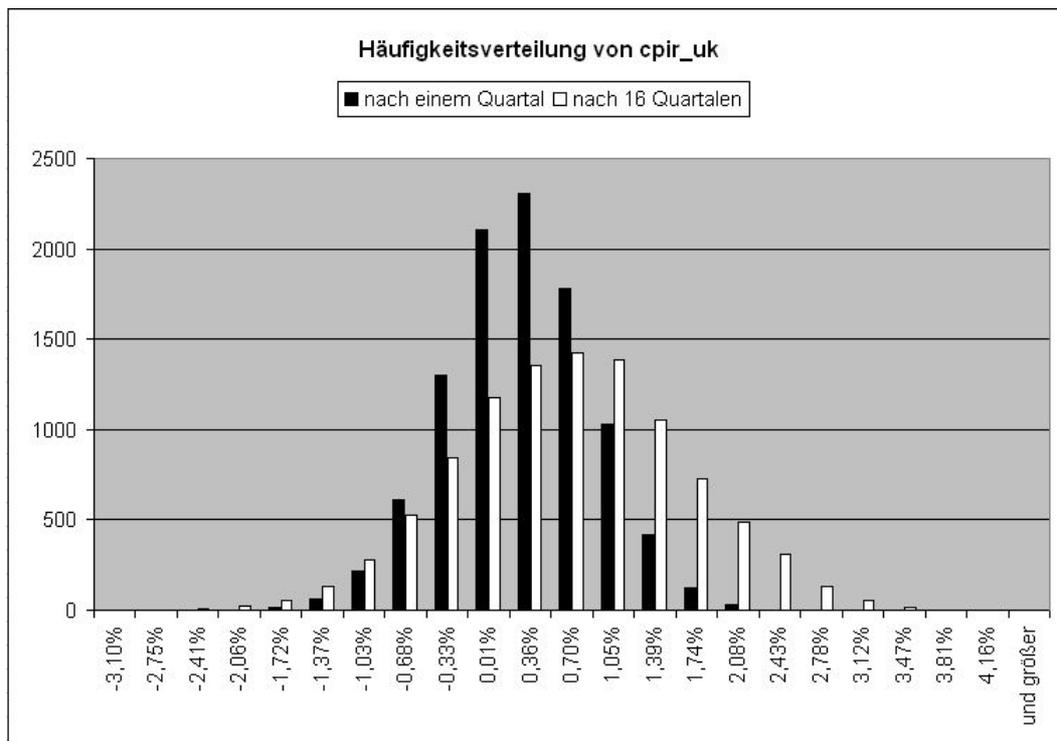


Abbildung 2

Die Abbildungen 3 und 4 zeigen den Einfluss des angewandten Verfahrens zur Simulation der Störgrößen nach 16 Quartalen. Die Normalverteilung betont die Ränder etwas stärker als das Bootstrap-Verfahren.

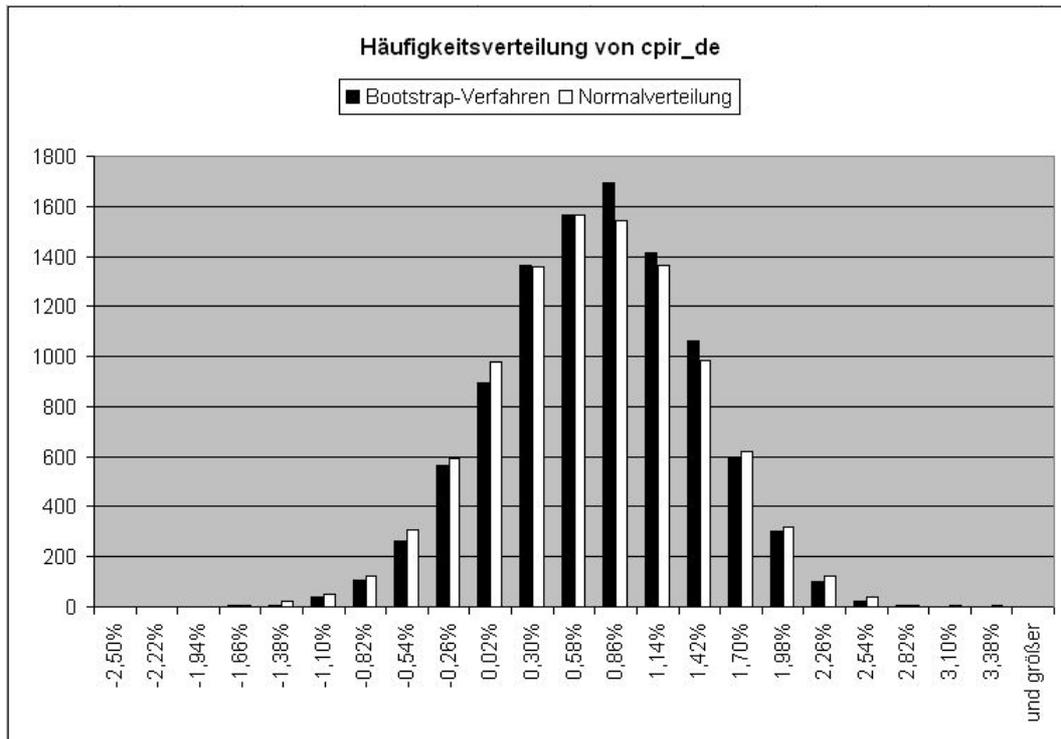


Abbildung 3

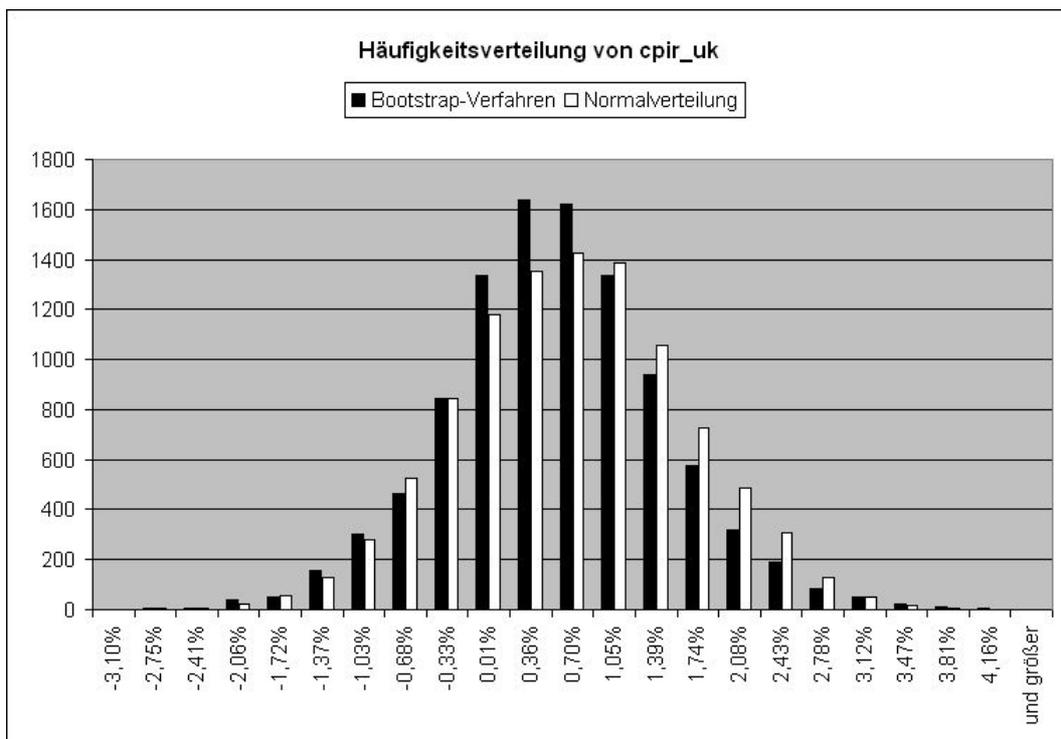


Abbildung 4

Nach 10.000 Simulationsläufen ergeben sich für das VAR-Modell folgende Werte (Störgrößen wurden mit dem Bootstrap-Verfahren simuliert):

**Tabelle 9** cpir\_de

	nach einem Quartal	nach 16 Quartalen
$\mu$	0,70 %	0,56 %
$\sigma$	0,36 %	0,64 %

**Tabelle 10** cpir\_uk

	nach einem Quartal	nach 16 Quartalen
$\mu$	0,11 %	0,41 %
$\sigma$	0,58 %	0,92 %

Beim Vergleich mit den Tabellen 7 und 8 stellt man nur kleine Abweichungen fest, was nicht verwundert, da die Regressionsgleichungen der univariaten Modelle gemäß den Tabellen 3 und 5 eine große Übereinstimmung mit dem VAR-Modell aufweisen.

**Literatur**

- Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N./Csáki (Hrsg.): Proceeding of the Second International Symposium on Information Theory, Akadémiai Kiadó, Budapest, S. 267-281.
- Allen, P. G./Fildes, R. (2007): Levels, differences and ECMs - Principles for Improved Econometric Forecasting. In: Oxford Bulletin of Economics and Statistics, erscheint demnächst, <http://www.umass.edu/resec/faculty/allen/obes.pdf>, Abruf am 2007-07-12
- Brüggemann, R./Lütkepohl, H.: Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System. In: Friedmann, R./Knüppel, L./Lütkepohl, H. (Hrsg.): Econometric Studies: A Festschrift in Honour of Joachim Frohn. LIT, Münster, S. 107-128.
- Chow, G. C. (1960): Tests of equality between sets of coefficients in two linear regressions. In: *Econometrica*, Vol. 28, S. 591-605.
- Christoffersen, P. F./Diebold, F. X. (1997): Cointegration and Long-Horizon Forecasting. NBER Working Paper No. T0217, <http://ssrn.com/abstract=226625>.
- Dickey, D. A./Fuller, W. A. (1979): Distribution of the Estimators for Autoregressive Time Series with a Unit Root. In: *Journal of the American Statistical Association*, Vol. 74, S. 427-431.
- Dickey, D. A./Fuller, W. A. (1981): Likelihood ratio statistics for autoregressive time series with a unit root. In: *Econometrica*, Vol. 49, S. 1057-1072.
- Doornik, J. A./Hansen, H. (1994): A practical test of multivariate normality. Unveröffentlichtes Manuskript, Nuffield College, University of Oxford.
- Embrechts, P./McNeil, A./Straumann, D. (2002): Correlation and dependence in risk management: properties and pitfalls. In: Dempster, M. A. H. (Hrsg.): *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, S. 176 - 223.
- Engle, R. F./Granger, C. W. J. (1987): Co-integration and Error Correction: Representation, Estimation and Testing. In: *Econometrica*, Vol. 55, S. 251-276.
- Engle, R. F./Yoo, B. S. (1987): Forecasting and testing in cointegrated systems. In: *Journal of Econometrics* Vol. 35, S. 143-159.
- Granger, C. W. J. (1981): Some properties of time series data and their use in econometric model specification. In: *Journal of Econometrics*, supplement 16, S. 121-130.
- Granger, C. W. J./Newbold, P. (1974): Spurious Regression in Econometrics. In: *Journal of Econometrics*, Vol. 2, S. 111-120.

- Krämer, W./Runde, R. (2000): Peaks or tails - What distinguishes financial data? In: Empirical Economics, Vol. 25, No. 4, S. 665-671.
- Ljung, G. M./Box, G. E. P.(1978): On a Measure of Lack of Fit in Time Series Models. In: Biometrika, Vol. 65, No. 2, S. 297-303.
- Lütkepohl, H. (2004): Vector Autoregressive and Vector Error Correction Models. In: Lütkepohl, H./Krätzig, M. (Hrsg.): Applied Time Series Econometrics, Cambridge University Press, Cambridge, S. 86 – 158.
- Lütkepohl, H. (2005): New Introduction to Multiple Time Series Analysis. Springer, Berlin, Heidelberg, New York.
- Schwarz, G. (1978): Estimating the Dimension of a Model. In: Annals of Statistics, Vol. 6, S. 461-464.
- Silverman, B. W. (2003): Density Estimation for Statistics an Data Analysis, Chapman-Hall, London.
- Stephens, M. A. (1974): EDF Statistics for Goodness of Fit and Some Comparisons. In: Journal of the American Statistical Association, Vol. 69, S. 730-737.
- Tiao, G. C./Tsay, R. S./Wang, T. (1993): Usefulness of Linear Transformations in Multivariate Time-Series Analysis. In: Empirical Economics, Vol. 18, S. 567-593.