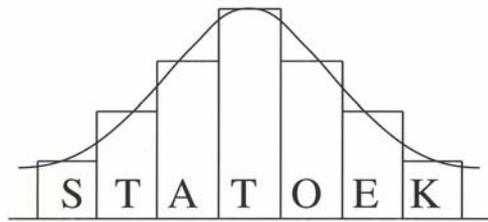


# Institut für Statistik und Ökonometrie



## **Die Verallgemeinerte Momentenmethode (Generalized Method of Moments - GMM) Darstellung und Anwendung**

Ralf Becker

Arbeitspapier Nr. 16 (Juni 1998)

## **Johannes Gutenberg-Universität Fachbereich Rechts- und Wirtschaftswissenschaften**

Haus Recht und Wirtschaft II  
D 55099 Mainz  
Tel: 0 61 31 - 39 22551 Fax: 0 61 31 - 39 23717  
E-Mail: STATOEK@Uni-Mainz.de

**Herausgeber: Univ.-Prof. Dr. Peter M. Schulze**

ISSN Nr. 1430 - 2136

Institut für Statistik und Ökonometrie  
Johannes Gutenberg-Universität Mainz  
Fachbereich Rechts- und Wirtschafts-  
wissenschaften  
Haus Recht und Wirtschaft II

D 55099 Mainz

Herausgeber: Univ.-Prof. Dr. P.M. Schulze

© 1998 Institut für Statistik und Ökonometrie, Mainz  
ISSN Nr. 1430 - 2136

## Inhaltsverzeichnis

Variablen- und Abkürzungsverzeichnis .....	I
1 Einleitung.....	1
2 Theoretischer Teil.....	1
2.1 Verallgemeinerte Momentenmethode (Generalized Method of Moments -GMM).....	1
2.1.1 Grundidee .....	2
2.1.2 Formale Darstellung der GMM.....	2
2.1.3 Auswahl der Momente und Instrumente .....	5
2.1.4 Verteilung der Schätzer und optimale Wahl der Gewichtungsmatrix.....	7
2.2 Schätzung der Varianz-Kovarianzmatrix $S$ .....	9
2.2.1 Unkorrelierte Residuen .....	9
2.2.2 Autokorrelierte Residuen .....	10
2.3 Testen im Rahmen der GMM .....	12
2.3.1 Parametertests.....	12
2.3.1.1 Signifikanz der Parameter .....	12
2.3.1.2 Restriktionen über mehrere Parameter.....	13
2.3.2 Überidentifizierende Restriktionen .....	15
2.3.3 Strukturbruch.....	16
2.4 Zwischenfazit zur GMM.....	18
3 Empirischer Teil - Schätzung des Zinsprozesses mittels GMM.....	19
3.1 Formale Darstellung der Schätzproblematik .....	20
3.2 Ergebnisse zum LEVEL-Modell .....	21
4 Fazit und Ausblick.....	24
Anhang .....	26
Literaturverzeichnis.....	30

## Abstract

In dieser Arbeit wird die Schätzmethode der Verallgemeinerten Momente (Generalized Method of Moments - GMM) vorgestellt. Neben der formalen Darstellung wird besonderes Augenmerk auf die notwendige Schätzung der Gewichtungsmatrix und auf Testmethoden im Rahmen der GMM gelegt. Anhand einer Schätzung des stochastischen Prozesses des kurzfristigen Zinssatzes wird die Anwendung der GMM verdeutlicht.

This paper presents the Generalized Method of Moments (GMM) as a technique of parameter estimation. Besides the general framework emphasis is put on the estimation of the weighting matrix and testing methods. The application of the GMM is illustrated by means of parameter estimation of the stochastic short term interest rate process.

## Variablen- und Abkürzungsverzeichnis

$\varepsilon$	Residuenvektor aus den Orthogonalitätsbedingungen $f$
$\theta$	Parametervektor
$f$	Vektor der Orthogonalitätsbedingung aus der Kombination von $h$ und dem Instrumentvektor
$g$	Stichprobenfunktion der Orthogonalitätsbedingungen $f$
$h$	Vektor der Momentbedingungen
$i$	1-Monats-Geldmarktsatz
$q$	Anzahl der zu schätzenden Parameter
$r$	Anzahl der Orthogonalitätsbedingungen
$t$	Zeitindex
$u$	Residuenvektor aus den Modell-Orthogonalitätsbedingungen
$x$	Vektor der erklärenden und der abhängigen Variablen zum Zeitpunkt $t$
$z$	Instrumentvektor
$A$	Koeffizientenmatrix aus der VAR-Schätzung
$\Lambda$	Jacobi-Matrix der partiellen Ableitungen der Parameterrestriktionen nach den Parametern
$F$	Jacobi-Matrix der partiellen Ableitungen der Orthogonalitätsbedingungen nach den Parametern
$J$	Zu minimierende Zielfunktion bei der GMM-Schätzung
$\Omega$	Informationsmenge
$S$	Varianz- Kovarianzmatrix der Orthogonalitätsbedingungen $f$
$T$	Stichprobenumfang
$V$	Varianz- Kovarianzmatrix der Parameterschätzer
$W$	Gewichtungsmatrix in der Zielfunktion $J$
$X$	$(T \times \text{Anzahl der Variablen})$ Matrix der beobachtbaren Variablen
AKF	Autokorrelationsfunktion
BLU	Best Linear Unbiased
GMM	Generalized Method of Moments
HAC	Heteroscedasticity and autocorrelation consistent
LM	Lagrange Multiplier
LR	Likelihood Ratio
ML	Maximum Likelihood
MSE	Mean Square Error
OLS	Ordinary Least Squares
PAKF	Partielle Autokorrelationsfunktion

## 1 Einleitung

Die Schätzmethode der Verallgemeinerten Momente (Generalized Method of Moments - GMM) erfreut sich, nachdem sie in ihrer allgemeinen Form 1982 von Lars Peter Hansen<sup>1</sup> eingeführt wurde, auf vielerlei Gebieten der empirischen Ökonomie großer Beliebtheit. Dies entspringt vor allem der Tatsache, daß diese Schätzmethode, unter sehr allgemeinen Bedingungen, konsistente Schätzer für Modellparameter hervorbringt. Unter speziellen Verteilungsannahmen bezüglich der verwendeten Daten bzw. der Residuen lassen sich andere Schätzmethoden finden, die entweder effizienter sind, oder sogar wünschenswerte Kleinstichprobeneigenschaften besitzen. Sind diese Verteilungsannahmen im konkreten Fall nicht gerechtfertigt, dann bietet die GMM die Möglichkeit, zumindest konsistente Schätzer für die Modellparameter abzuleiten.

Angewandt werden soll die GMM im Rahmen dieser Arbeit auf den stochastischen Prozeß des kurzfristigen, risikolosen Zinssatzes, dessen Modellierung im Rahmen von Zinsstrukturmodellen notwendig ist. Mit Hilfe solcher Modelle können zinsderivate Finanzprodukte bewertet werden.

Im Anschluß an diese Einleitung wird im theoretischen Teil der Arbeit die GMM als Schätzmethode vorgestellt. Besonderes Augenmerk wird dabei auf die Schätzung der Gewichtungsmatrix und die Testmethoden bei der GMM gelegt. In Kapitel 3 folgt eine empirische Analyse des stochastischen Zinsprozesses. Die Arbeit schließt mit einem Fazit.

## 2 Theoretischer Teil

### 2.1 Verallgemeinerte Momentenmethode (Generalized Method of Moments - GMM)

Die Methode der GMM schließt eine ganze Reihe anderer Schätzmethoden als Spezialfälle ein. So lassen sich auch die Kleinstquadrat- oder die Maximum Likelihood (ML) Schätzer als GMM-Schätzer beschreiben.<sup>2</sup> Insofern ist in der folgenden Arbeit die Rede von einem Spezialfall aus der Klasse der GMM-Schätzer, dem *Generalized Instrumental Variable Estimator*, dem sowohl in der theoretischen wie auch in der empirischen Literatur die größte Bedeutung zukommt.

Beim später behandelten Schätzproblem zum stochastischen Zinsprozeß handelt es sich um eine Spezifikation, die nichtlinear in ihren Parametern ist. Zwar gibt es für diese Klasse von Modellen auch alternative Schätzmethoden, wie die nichtlineare

---

<sup>1</sup> Vgl. Hansen, 1982.

<sup>2</sup> Vgl. Ogaki, 1993a, S.458.

Kleinstquadratmethode oder die ML-Methode, doch ist es für beide Methoden notwendig, zusätzliche Annahmen bezüglich der Verteilung der Residuen bzw. der abhängigen Variablen zu treffen. Wird ein Modell mit Hilfe dieser Verfahren geschätzt, dann testet man gleichzeitig die Gültigkeit dieser Annahmen.<sup>3</sup> Die GMM benötigt hingegen keinerlei Verteilungsannahmen bezüglich der Residuen. Auch wenn die Residuen bedingte Heteroskedastizität oder serielle Autokorrelation aufweisen, bleiben die geschätzten Parameter asymptotisch konsistent.<sup>4</sup>

### 2.1.1 Grundidee

Die Grundidee der GMM besteht darin, ein Modell zu finden, welches einen Vektor der beobachtbaren, abhängigen Variablen  $y$  möglichst gut durch die Matrix  $X$ , der unabhängigen Variablen, erklärt. Zu schätzen ist der Parametervektor  $\theta$  in  $y = f(X, \theta)$ . Das Maß welches den GMM-Schätzer für  $\theta$  selektiert unterscheidet die nichtlineare Kleinstquadratmethode oder die ML-Methode von der GMM und gibt letzteren den Namen. Während die Kleinstquadratmethode die quadrierten Residuen minimiert und die ML-Methoden die Log-Likelihood-Funktion maximiert, minimiert die GMM die Differenz zwischen beobachteten Momenten der abhängigen Variablen und den aus dem geschätzten Modell erwarteten Momenten. Die Modellparameter werden so gewählt, daß die erwarteten Momente aus dem Modell möglichst genau den beobachteten entsprechen. Für jedes Moment wird eine Momentbedingung aufgestellt, die es zu minimieren gilt.<sup>5</sup>

### 2.1.2 Formale Darstellung der GMM

Um die formale Einführung der GMM möglichst einfach und übersichtlich zu gestalten, wird in diesem Kapitel davon ausgegangen, daß eine Momentbedingung aufgestellt wird. Dies sei z.B. eine Formulierung für den Erwartungswert der Zinsänderung, die in der später behandelten Anwendung  $E(\Delta i_t) = \alpha + \beta i_{t-1}$  lautet.

$X$  sei im folgenden die Matrix der beobachtbaren abhängigen und unabhängigen Variablen und  $x_t$  deren Vektor zum Zeitpunkt  $t$ .<sup>6</sup>  $\theta_0$  sei der unbekannte, wahre Parametervektor,  $h$  der funktionale Zusammenhang (Moment) und  $u_t$  der resultierende Residuenvektor. Damit läßt sich die Momentbedingung als Erwartungswert von  $u_t$  zum Zeitpunkt  $t$  mit der zu diesem Zeitpunkt zur Verfügung stehenden Informationsmenge  $\Omega_t$ ,

---

<sup>3</sup> Vgl. Haverkamp, 1993, S.70.

<sup>4</sup> Vgl. Hansen, 1982, S.1030.

<sup>5</sup> Vgl. Hansen/Singleton, 1982, S.1274.

<sup>6</sup> Dies können auch verzögerte Variablen sein, die zum Zeitpunkt  $t$  zur Verfügung stehen.

$E(u_t|\Omega_t) = 0$  mit  $u_t = h(x_t, \theta_0) = \Delta i_t - \alpha - \beta i_{t-1}$ , beschreiben.<sup>7 8</sup> Es ist zu gewährleisten, daß  $E(u_t) = 0$  nur für den wahren Parametervektor  $\theta_0$  gilt und es keinen zweiten Parametervektor  $\tilde{\theta}$  gibt, der per se  $u_t = 0 \forall x_t$  setzt.<sup>9</sup>

Besteht der Parametervektor  $\theta$  aus  $q$  Elementen, dann benötigt man mindestens ebenso viele Momentbedingungen, um die  $q$  Parameter zu bestimmen. Sind aus dem theoretischen Modell keine weiteren Momentbedingungen abzuleiten, macht man sich zunutze, daß  $E(u_t|\Omega_t) = 0$  eine bedingte Momentbedingung ist, und der Erwartungsfehler  $u_t$  mit Instrumenten  $z_t$  aus  $\Omega_t$  unkorreliert ist. Somit gilt  $E(u_t z_t) = 0$ .<sup>10</sup>

Der  $(r \times 1)$  Vektor  $z_t$  sei der Vektor der zum Zeitpunkt  $t$  beobachteten Instrumente. Aus der Operation

$$f(x_t, \theta, z_t) = h(x_t, \theta) z_t = \left( h(x_t, \theta) \quad h(x_t, \theta) z_{1,t} \quad \dots \quad h(x_t, \theta) z_{r-1,t} \right)' \quad (1)$$

folgt der  $(r \times 1)$  Vektor der Orthogonalitätsbedingungen.<sup>11</sup> Um die  $q$  Parameter schätzen zu können, muß  $r \geq q$  gelten. In der Regel ist das erste Element von  $z_t$  die Konstante 1. Somit bleibt die aus dem Modell abgeleitete Momentbedingung als erstes Element von  $f$  erhalten.

Für den wahren Parametervektor  $\theta_0$  gilt  $E[f(x_t, \theta_0, z_t)] = 0$ . Mit  $r = q$ , d.h. genauso vielen Instrumenten wie gesuchten Parametern, findet man den Schätzer für  $\theta$ , indem man die Stichprobenfunktion

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^T f(x_t, \theta, z_t) = \frac{1}{T} \sum_{t=1}^T \varepsilon_t$$

gleich Null setzt und das resultierende Gleichungssystem mit  $r$  Gleichungen nach den unbekanntem Parametern löst. Wegen  $r = q$  ist dieses Gleichungssystem genau identifiziert. Für  $r > q$  wird  $g_T(\theta)$  i.d.R. ungleich Null sein. In diesem Fall wird die Zielfunktion

$$J_T(\theta) = g_T(\theta)' W_T g_T(\theta) \quad (2)$$

$\begin{matrix} (1 \times 1) & & & \\ & (1 \times r) & (r \times r) & (r \times 1) \end{matrix}$

<sup>7</sup> Vgl. Hansen/Singleton, 1982 S.1274-75.

<sup>8</sup> Die Menge der Variablen, die zum Zeitpunkt der Erwartungsbildung zur Verfügung stehen, nennt man Informationsmenge. Vgl. Davidson, MacKinnon, 1993, S.52.

<sup>9</sup> Vgl. Ogaki, 1993a, S.464.

<sup>10</sup> Vgl. Davidson/MacKinnon, 1993, S.584.

<sup>11</sup> Mit einer aus dem Modell abgeleiteten Momentbedingung ist  $h$  ein Skalar. Benutzt man mehrere solcher Momentbedingungen wird  $h$  zu einem Vektor und es ist das Kronecker Produkt aus  $h$  und  $z$  zu bilden. Im Folgenden werden direkt aus dem theoretischen Modell abgeleitete Bedingungen als Momentbedingungen bezeichnet, auch wenn diese streng genommen keine Momente sein müssen (vgl. Abschnitt 2.1.3). Bedingungen, die aus der Kombination mit einem Instrumentvektor entstehen, werden Orthogonalitätsbedingungen genannt.

minimiert. Die Matrix  $W_T$  sei eine  $(r \times r)$  Gewichtungsmatrix.<sup>12</sup> Der GMM-Schätzer  $\hat{\theta}_T$  ergibt sich aus der Lösung der Minimierungsbedingung der Gleichung (2),

$$\Gamma(\hat{\theta}_T, x_t, z_t)' W_T g(\hat{\theta}_T, x_t, z_t) = 0 \quad (3)$$

$(q \times r)$                        $(r \times r)$                        $(r \times 1)$

mit

$$\Gamma(\hat{\theta}_T, x_t, z_t)' = \left\{ \frac{\partial g(\theta; x_t, z_t)}{\partial \theta'} \Big|_{\theta = \hat{\theta}_T} \right\}$$

als der  $(q \times r)$  Jacobi-Matrix der partiellen Ableitungen der Orthogonalitätsbedingungen nach den zu schätzenden Parametern, bewertet an der Stelle  $\hat{\theta}_T$ . In Gleichung (3) werden aus den ursprünglich  $r$  Orthogonalitätsbedingungen  $q$  Linearkombinationen gebildet. Mit welcher Gewichtung die Orthogonalitätsbedingungen in die Linearkombinationen eingehen, wird durch die Matrix  $W_T$  bestimmt. Es resultiert ein Gleichungssystem mit  $q$  Gleichungen und ebenso vielen Unbekannten, den zu schätzenden Parametern.<sup>13</sup> Hansen und Singleton nennen den resultierenden Schätzer für den Parametervektor  $\theta$ , wenn er mit Hilfe der Instrumentvariablen abgeleitet wird, auch einen *Generalized Instrumental Variable Estimator*.<sup>14</sup>

Problematisch ist nun, daß man zur Herleitung des GMM-Schätzers für  $\theta$  die Gewichtungsmatrix  $W_T$  benötigt, diese aber wiederum aus den geschätzten Residuen geschätzt werden muß.<sup>15</sup> In Abschnitt 2.2 wird gezeigt, daß sich die optimale Wahl für  $W_T$  aus der geschätzten Varianz-Kovarianzmatrix  $S_T$  der Orthogonalitätsbedingungen  $f(x_t, \theta, z_t)$  ergibt. Um die aus  $f$  resultierenden Residuen zu schätzen, benötigt man einen geschätzten Parametervektor  $\hat{\theta}_T$ . Um unter dieser zirkulären Bedingung einen konsistenten Schätzer für  $\theta$  zu erhalten, wendet man die folgende, iterative Schätzstrategie an.

Einen ersten Parametervektor  $-\hat{\theta}_T^{(1)}$  erhält man, indem man in Gleichung (2) die Gewichtungsmatrix  $W_T$  durch die  $(r \times r)$  Einheitsmatrix  $I_r$  ersetzt und (2) dann minimiert. Dieses Ergebnis kann zu einer ersten Approximation an  $S$ , der Varianz-Kovarianzmatrix von  $f$  -  $\hat{S}_T^{(1)}$  - genutzt werden. Mit  $\hat{W}_T^{(1)} = (\hat{S}_T^{(1)})^{-1}$  als neuer Gewichtungsmatrix in (2) kann dann ein neuer Schätzer für den Parametervektor  $\theta$  -  $\hat{\theta}_T^{(2)}$  - geschätzt werden.

<sup>12</sup> Im Abschnitt 2.2 wird gezeigt, wie diese Matrix aus den Daten geschätzt wird.

<sup>13</sup> Vgl. Hamilton, 1994, S.414.

<sup>14</sup> Vgl. Hansen/Singleton, 1982, S.1276.

<sup>15</sup> Vgl. Hamilton, 1994, S.413.

1.  $Min J_T = g_T' I_r g_T \rightarrow \hat{\theta}_T^{(1)}$
2. Berechne  $\hat{S}_T^{(1)}(\hat{\theta}_T^{(1)}) \rightarrow \hat{W}_T^{(1)}$
3.  $Min J_T = g_T' \hat{W}_T^{(1)} g_T \rightarrow \hat{\theta}_T^{(2)}$
4. Berechne  $\hat{S}_T^{(2)}(\hat{\theta}_T^{(2)}) \rightarrow \hat{W}_T^{(2)}$

usw.

Diese Iteration kann beliebig oft durchgeführt werden, bis die Unterschiede in zwei aufeinanderfolgenden Parameterschätzungen so klein sind, daß ein diesbezügliches Abbruchkriterium greift. Allerdings ist auch der geschätzte Vektor  $\hat{\theta}_T^{(1)}$  schon asymptotisch konsistent. Weitere Iterationen bringen dennoch den Vorteil, daß die geschätzten Parameter unabhängig von der gewählten Gewichtungsmatrix  $W_0$  sind.<sup>16</sup>

### 2.1.3 Auswahl der Momente und Instrumente

Ein Vorteil der GMM ist, daß, um Modellparameter zu schätzen, nicht das gesamte Modell formuliert werden muß, sondern auf sogenannte Momentbedingungen zurückgegriffen werden kann. Weiß man, daß eine Variable  $x_t$  einer bestimmten Verteilung genügt, dann lassen sich z.B. Aussagen über die ersten beiden Momente, den Erwartungswert und die Varianz dieser Variable ableiten. Ist  $x_t$  beispielsweise exponentialverteilt, mit der Dichtefunktion

$$j(x|\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \text{ mit } \lambda > 0 \\ 0 & \text{sonst} \end{cases},$$

dann folgt, daß  $E(x) = 1/\lambda$  und  $VAR(x) = 1/\lambda^2$  gilt. Der Verteilungsparameter  $\lambda$  ließe sich nun entweder unter Verwendung der Dichtefunktion  $j$  und der ML-Methode, oder unter Zuhilfenahme des  $(2 \times 1)$  Stichprobenvektors

$$g(x_t, \lambda) = \begin{pmatrix} \frac{1}{T} \sum x - \frac{1}{\lambda} \\ \frac{1}{T} \sum (x - \bar{x})^2 - \frac{1}{\lambda^2} \end{pmatrix}$$

und der GMM schätzen. Ist die Verteilung bekannt, kann  $\lambda$  mit der ML effizient geschätzt werden, und man würde nur dann auf die GMM zurückgreifen, wenn das Erstellen der Log-Likelihood Funktion für die ML Schätzung zu aufwendig ist.

Im später behandelten Beispiel werden Parameter des stochastischen Zinsprozesses geschätzt. Die Prozeßgleichung wird zwar als bekannt vorausgesetzt, doch implizieren verschiedene Werte einzelner Parameter unterschiedliche Verteilungen für den Zinssatz. Die Verteilung kann demnach nicht a priori festgelegt werden, was allerdings für eine ML-Schätzung unerlässlich ist. ML-Schätzer sind nur unter der Gültigkeit einer

expliziten Verteilungsannahme konsistent und effizient. Benutzt man die aus dem Modell abgeleiteten Modellbedingungen zu den ersten beiden Momenten und schätzt die Parameter mit Hilfe der GMM, so kann man dieses Problem vermeiden.

Häufig lassen sich Orthogonalitätsbedingungen auch in Form von Optimierungskalkülen ableiten. Das klassische Beispiel hierzu sind Anwendungen aus dem Gebiet der Theorie der rationalen Erwartungen. Hansen und Singleton leiten aus einem *consumption based capital asset pricing model* (C-CAPM) das Optimierungskalkül eines repräsentativen Agenten her. Zusammen mit der Annahme, daß Erwartungen rational gebildet werden, d.h. daß Erwartungsfehler aus diesem Kalkül mit sämtlichen, zum Zeitpunkt der Erwartungsbildung verfügbaren Variablen unkorreliert sind, lassen sich Orthogonalitätsbedingungen ableiten.<sup>17</sup>

Kann man hieraus nicht mindestens so viele Momentbedingungen ableiten wie Parameter zu schätzen sind, so greift man auf die Annahme zurück, daß die Residuen aus den Momentbedingungen mit Variablen aus der Informationsmenge  $\Omega$  unkorreliert sind. Gemäß Gleichung (1) können daraus weitere Orthogonalitätsbedingungen konstruiert werden.

Hamilton beschreibt die Anforderungen, die an die benutzten Instrumente gestellt werden müssen.<sup>18</sup> Die Instrumentvariablen müssen aus dem Informationsset  $\Omega_t$  zum Zeitpunkt  $t$  stammen, dürfen folglich keine zukünftigen Variablen sein, deren Werte heute noch nicht bekannt sind. Sie sollten außerdem mit den Residuen aus den theoretisch abgeleiteten Momentbedingungen unkorreliert sein. Dies ist die Bedingung dafür, daß sich gemäß Gleichung (1) weitere Orthogonalitätsbedingungen ableiten lassen.<sup>19</sup>

Hansen und Singleton schlagen vor, zeitverzögerte Variablen aus  $X$  als Instrumente zu benutzen.<sup>20</sup> Unter diesen lassen sich i.d.R. auch solche Instrumente finden, die mit den erklärenden Modellvariablen hochkorreliert sind.<sup>21</sup>

Die Auswahl der Momentbedingungen und der Instrumente ist neben der Schätzung der optimalen Gewichtungsmatrix sicherlich das Kernproblem bei der Anwendung der GMM. Leider liefert die Theorie kaum Hinweise zur optimalen Wahl der Orthogonalitätsbedingungen. Erst kürzlich wurden Verfahren zur Generierung von Momentbedingungen entwickelt, mit Hilfe derer Parameter auch effizient geschätzt werden kön-

---

<sup>16</sup> Vgl. Hall, 1993, S.403.

<sup>17</sup> Vgl. Hansen/Singleton, 1982, S.1278ff.

<sup>18</sup> Vgl. Hamilton, 1994, S.426f.

<sup>19</sup> Vgl. Ogaki, 1993a, S.479.

<sup>20</sup> Vgl. Hansen/Singleton, 1982, S.1281 und Tauchen, 1986, S.406.

<sup>21</sup> Vgl. Hamilton, 1994, S.426-427.

nen. Diese Methoden sind für eine Anwendung im Rahmen dieser Arbeit allerdings zu komplex und aufwendig.<sup>22</sup>

Wie sensibel die GMM auf die Wahl der Instrumente reagiert, beschreiben Ghysels und Hall, die feststellen, daß die Parameterschätzer z.T. nicht robust bzgl. der Wahl der Instrumente sind.<sup>23</sup>

### 2.1.4 Verteilung der Schätzer und optimale Wahl der Gewichtungsmatrix

Nachdem das Vorgehen bei einer GMM-Schätzung dargestellt wurde, gilt es, die resultierende Verteilung der Parameterschätzer zu bestimmen, also die Eigenschaften der GMM-Schätzer abzuleiten. Mit diesem Problem verbunden ist die Frage nach der optimalen Bestimmung der Gewichtungsmatrix  $W_T$ .

Bzgl. der Residuen wurde keine explizite Verteilungsannahme getroffen. Läßt sich der Zentrale Grenzwertsatz anwenden,<sup>24</sup> dann ist  $(1/\sqrt{T})\sum_{t=1}^T \varepsilon_t = \sqrt{T}g_T$  in großen Stichproben asymptotisch normalverteilt mit dem Mittelwert 0 und der Varianz-Kovarianzmatrix<sup>25</sup>

$$S = \lim_{j \rightarrow \infty} \sum_{-j}^j E(\varepsilon_t \varepsilon'_{t-j}) .$$

$E(\varepsilon_t \varepsilon'_{t-j})$  ist die Autokovarianzmatrix der Residuen  $\varepsilon_t$  und  $\varepsilon_{t-j}$  bewertet zum wahren Parameter  $\theta_0$ .  $S$  ist die asymptotische Varianz-Kovarianzmatrix des Stichprobenvektors  $f(x_t, \theta_0, z_t)$ .

Es läßt sich zeigen, daß unter diesen Bedingungen  $\sqrt{T}(\hat{\theta}_T - \theta_0)$  in großen Stichproben approximativ normalverteilt ist,  $\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N(0, V)$ . D.h., der GMM-Schätzer ist asymptotisch konsistent und normalverteilt. Mit einem Stichprobenumfang von  $T \rightarrow \infty$  strebt der GMM-Schätzer gegen seinen wahren Parameterwert. Hansen beschreibt die diesem Ergebnis zu Grunde liegenden Annahmen. Neben einigen technischen Voraussetzungen gehört dazu die Annahme, daß die Zeitreihen in der Variablenmatrix  $X$  stationär und ergodisch sind.<sup>26</sup> Außerdem müssen die Gewichtungsmatrix  $W$  sowie die Jacobi-Matrix  $\Gamma$  den vollen Rang besitzen.<sup>27</sup>

<sup>22</sup> Vgl. Gallant/Tauchen, 1996 und Hansen/Scheinkman, 1995.

<sup>23</sup> Vgl. Ghysels/Hall, 1990, S.134.

<sup>24</sup> Schervish, 1995, S.642: „Suppose that  $\{X_i\}_{i=1}^{\infty}$  is a sequence that is iid with finite mean  $\mu$  and finite variance  $\sigma^2$ . Let  $\bar{X}_n$  be the average of the first  $n$   $X_i$ s. Then  $\sqrt{n}(\bar{X}_n - \mu) \rightarrow N(0, \sigma^2)$ , the normal distribution with mean 0 and variance  $\sigma^2$ .“

<sup>25</sup> Vgl. Ogaki, M., 1993a, S.457.

<sup>26</sup> Ist eine Zeitreihe stationär und ergodisch, dann kann aus einer Realisation über einen endlichen Zeitraum auf den Prozeß der Zeitreihe geschlossen werden, da  $E[f(X_t)]_{T_0}^T = E[f(X_t)]_{T_0+\tau}^{T+\tau}$ . Speziell heißt das, daß sich die Stichprobenmomente, wenn der Stichprobenumfang gegen unendlich

Die Varianz  $V$  ist von der Wahl der Gewichtungsmatrix  $W$  abhängig und lautet unter diesen Voraussetzungen  $V = Cov(W_0) = (\Gamma'W_0\Gamma)^{-1}(\Gamma'W_0SW_0'\Gamma)(\Gamma'W_0\Gamma)^{-1}$ .<sup>28</sup> Die Wahl der Gewichtungsmatrix hat zwar keinen Einfluß auf die Konsistenzeigenschaft des Schätzers, doch ist sie entscheidend für deren Effizienz. Es wird der Schätzer gesucht, der die geringste Varianz aufweist. Diese ergibt sich aus den Diagonalelementen von  $Cov(W_0)$ . Hansen weist nach, daß der Parametervektor  $\theta$  dann asymptotisch effizient geschätzt wird, wenn  $W_0 = S^{-1}$  gewählt wird.  $V = Cov(W_0) = Cov(S^{-1})$  vereinfacht sich dann zu

$$V = Cov(S^{-1}) = (\Gamma'S^{-1}\Gamma)^{-1}. \quad (4)$$

Da sich die Matrix  $S$  aus den geschätzten Orthogonalitätsbedingungen  $f(x_i, \theta_0, z_i)$  berechnet, wird offensichtlich, daß die Wahl und die Anzahl der Instrumente Einfluß auf die Verteilung der geschätzten Parameter nimmt.

Wie Tauchen in einer Simulationsstudie zu den Kleinstichprobeneigenschaften der GMM-Schätzer feststellt, hat vor allem die Anzahl der Orthogonalitätsbedingungen einen Einfluß auf die Varianz der Schätzer und auf die Verzerrung, die in kleinen Stichproben auftritt.<sup>29</sup> Er spricht von einem „Variance/Bias Trade-off“. Er weist im Rahmen seiner Simulation nach, daß zusätzliche Orthogonalitätsbedingungen zwar zu einer kleineren Varianz, aber auch zu größeren Verzerrungen der Schätzer führen. Vor allem, wenn aus den geschätzten Parametern Konfidenzintervalle berechnet werden, ist dies eine problematische Konstellation, was Tauchen dazu veranlaßt, für einen gemäßigten Umgang mit überidentifizierenden Restriktionen<sup>30</sup> zu plädieren. Andersen und Sørensen bestätigen, daß bei kleinen Stichproben zu viele Momentrestriktionen suboptimal sind. Die Ursache hierzu liegt in der Tatsache, daß unter zusätzlichen Instrumenten die Präzision der Schätzung der optimalen Gewichtungsmatrix  $S^{-1}$  leidet. Wird die Anzahl der Orthogonalitätsbedingungen erhöht, steigt auch die Anzahl der zu schätzenden Elemente dieser Matrix.<sup>31</sup>

Es erscheint allerdings auch nicht sinnvoll, sich auf ein genau identifiziertes Modell zu verlassen, d.h. ein Modell, in dem die Anzahl der Orthogonalitätsbedingungen gleich der Anzahl der Parameter ist. Zusätzliche Orthogonalitätsbedingungen verkleinern die

---

strebt, mit der Wahrscheinlichkeit 1 den wahren Momenten annähern. Vgl. Nerlove et al., 1995, S.26-27.

<sup>27</sup> Vgl. Hansen, 1982, S.1039-1042.

<sup>28</sup> Vgl. Hall, 1993, S.401-402.

<sup>29</sup> Vgl. Tauchen, 1986, S.410-415. Er benutzt Stichprobenumfänge von 50 und 75.

<sup>30</sup> Schätzt man mit vier Orthogonalitätsbedingungen drei Parameter, dann besitzt die Schätzung eine überidentifizierende Restriktion.

<sup>31</sup> Vgl. Andersen/Sørensen, 1996, S.349 und S.334. Auch sie benutzen eine Kombination von Bias und Varianz als Kriterium. Während Tauchen bis zu fünf Orthogonalitätsbedingungen testet, überprüfen Andersen und Sørensen bis zu 24.

Varianz der GMM-Schätzer und verringern die Wahrscheinlichkeit mit der eine Konstellation auftritt, in der das nichtlineare Schätzverfahren zu keiner konvergierenden Lösung gelangt.<sup>32</sup> Außerdem ermöglichen überidentifizierende Restriktionen einen Test auf Mißspezifikation des Modells der in Abschnitt 2.3.2 vorgestellt wird.

## 2.2 Schätzung der Varianz-Kovarianzmatrix $S$

Neben den oben geschilderten Überlegungen zu der Auswahl der Instrumente, kommt der Wahl des geeigneten Schätzers für die Varianz-Kovarianzmatrix  $S$  eine besondere Bedeutung zu. Sie geht in die Schätzung des effizienten GMM-Schätzers ein und ist Bestandteil der Formel zur Berechnung der Kovarianzmatrix der Schätzer. Ebenso ist sie Bestandteil praktisch aller Prüfgrößen, die im Zusammenhang mit Hypothesentests von Relevanz sind.

### 2.2.1 Unkorrelierte Residuen

Sind die Residuen  $\varepsilon_t$  der  $r$  Orthogonalitätsbedingungen nicht autokorreliert, dann läßt sich die Varianz-Kovarianzmatrix  $S_T$  aus

$$\hat{S}_T^W = \frac{1}{T} \sum_{t=1}^T f(x_t, \hat{\theta}_T, z_t) f(x_t, \hat{\theta}_T, z_t)' = \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \quad (5)$$

schätzen.<sup>33</sup> Ist  $\hat{\theta}_T$  ein konsistenter Schätzer für  $\theta_0$ , also z.B.  $\hat{\theta}_T^{(1)}$ , dann ist  $\hat{S}_T$  ein konsistenter Schätzer für  $S$ . Dies ist der *White*-Schätzer für den Fall nicht autokorrelierter Residuen. Für den Fall, daß  $f$  aus drei Orthogonalitätsbedingungen besteht,  $f_t = (\varepsilon_{1t} \ \varepsilon_{2t} \ \varepsilon_{3t})$ , läßt sich dieser wie folgt berechnen:<sup>34</sup>

$$\hat{S}_T^W = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{pmatrix} (\varepsilon_{1t} \ \varepsilon_{2t} \ \varepsilon_{3t}) = \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varepsilon_{1t}^2 & & \\ \varepsilon_{1t} \varepsilon_{2t} & \varepsilon_{2t}^2 & \\ \varepsilon_{1t} \varepsilon_{3t} & \varepsilon_{2t} \varepsilon_{3t} & \varepsilon_{3t}^2 \end{pmatrix}.$$

Auf der Diagonalen von  $\hat{S}_T^W$  stehen die Varianzen der Residuen aus den Orthogonalitätsbedingungen. Als Gewichtungsmatrix bei der GMM-Schätzung dient dann  $W = \hat{S}_T^{W^{-1}}$ . Ist z.B. das Element (1,1), d.h. die Varianz der ersten Orthogonalitätsbedingung, relativ groß, wird das Element (1,1) der Gewichtungsmatrix  $W$  relativ klein sein. Dies bedeutet, daß Orthogonalitätsbedingungen mit großer (geringer) Varianz, d.h. solche, die vom Modell unpräzise (präzise) erklärt werden, ein kleineres (größeres) Gewicht erhalten.

<sup>32</sup> Vgl. Andersen/Sørensen, 1996, insbesondere Tabellen auf den Seiten 333, 337 und 340.

<sup>33</sup> Vgl. Hamilton, 1994, S.413.

<sup>34</sup> Symmetrische Matrizen werden nur als Dreiecksmatrix dargestellt.

## 2.2.2 Autokorrelierte Residuen

In der Autokorrelation der Residuen stecken Informationen, deren Nutzung zu verbesserten Schätzern führt. Die Vernachlässigung dieser Informationen kann dazu führen, daß ein eigentlich korrekt spezifiziertes Modell als fehlspezifiziert getestet wird (vgl. Abschnitt 2.3.2). Dies wird vermieden, indem ein *heteroscedasticity and autocorrelation consistent* (HAC) Schätzer für  $S$  zur Anwendung kommt. Ein solcher HAC Schätzer soll im folgenden beschrieben werden.

Sind die Residuen der Orthogonalitätsbedingungen autokorreliert, gilt es, die Varianz-Kovarianzmatrix der Grundgesamtheit

$$S = \Pi_0 + \sum_{v=1}^{\infty} (\Pi_v + \Pi_v') \text{ mit } \Pi_v \equiv E(f_t f_{t-v}')$$

zu schätzen.<sup>35</sup> Wird  $S$  dadurch geschätzt, daß lediglich die wahren, aber unbekanntenen Autokovarianzen  $\Pi_v$  (bis zu einem Lag  $m$ ) durch deren Schätzwerte gemäß Gleichung (5) ersetzt werden, gelangt man zum *White* Schätzer für autokorrelierte Residuen. Der *White* Schätzer ist allerdings bei kleinen Stichproben nicht notwendigerweise positiv semi-definit. Dies kann u.U. zu negativen Varianzen führen, wodurch Hypothesentests unmöglich werden. Auch beim iterativen Vorgehen bei der GMM-Schätzung können Probleme auftreten, wenn  $S$ , dessen Inverse als Gewichtungsmatrix benutzt wird, nicht positiv semi-definit ist. Dieses Problem kann mit HAC Schätzern vermieden werden.<sup>36</sup> Bei der Auswahl des Schätzers gilt es drei Entscheidungen zu treffen.

1. Welches Fenster (*kernel*)<sup>37</sup> wird benutzt?
2. Wieviel Lags werden berücksichtigt, bzw. wie groß ist die Bandbreite  $m$  des Fensters?
3. Werden die Residuen einer Transformation (*Pre-Whitening*) unterzogen, die einen Teil der autoregressiven Struktur der Restwerte beseitigt?

Im folgenden soll kurz auf diese drei Problemstellungen eingegangen werden. Mit verschiedenen *kernels* erhält man Schätzer für  $S$ , die sich in ihrem Bias und ihrer Varianz unterscheiden. Simulationsstudien müssen hier über die optimale Wahl in bestimmten Situationen entscheiden.<sup>38</sup> Im Rahmen dieser Arbeit können die verschiedenen Ausprägungen dieser Schätzer nicht ausführlich behandelt werden. Andersen und Sørensen kommen zu dem Ergebnis, daß der Newey-West-Schätzer die besten Eigen-

---

<sup>35</sup> Dies folgt aus  $S = \sum_{v=-\infty}^{\infty} \Pi_v$  und  $\Pi_{-v} = \Pi_v'$ . Vgl. Davidson/MacKinnon, 1993, S.609-610.

<sup>36</sup> Vgl. Chang, 1984, S.320. Zu den ökonomischen Folgen siehe Hamilton, 1994, S.281 und Newey/West, 1987a, S.704.

<sup>37</sup> Vereinfacht gesagt ist dies die Gewichtungsvorschrift für die Autokovarianzen  $v$ -ter Ordnung. Vgl. Hartung, 1986, S.709-714.

<sup>38</sup> Vgl. Ogaki, 1993a, S.466.

schaften besitzt.<sup>39</sup> Unter bestimmten Bedingungen sind zwar andere Fenster zu bevorzugen, doch rechtfertigen die eher marginalen Verbesserungen die wesentlich erhöhte Komplexität i.d.R. nicht, so daß in dieser Arbeit nur auf den Newey-West-Schätzer zurückgegriffen wird.<sup>40</sup>

Der Newey-West-Schätzer für  $S$  wird wie folgt berechnet:<sup>41</sup>

$$\hat{S}_T^{NW} = \hat{\Pi}_{0,T} + \sum_{v=1}^m (1 - (v/(m+1))) (\hat{\Pi}_{v,T} + \hat{\Pi}_{v,T}'),$$

$$\text{wobei } \hat{\Pi}_{v,T} = \frac{1}{T} \sum_{t=v+1}^T f(x_t, \hat{\theta}_T, z_t) f(x_{t-v}, \hat{\theta}_T, z_{t-v})' = \frac{1}{T} \sum_{t=v+1}^T \varepsilon_t \varepsilon_{t-v}'.$$

Für  $v = 0$  ( $v$  gibt die Ordnung der Autokorrelation an) wird  $\hat{\Pi}_{v,T}$  zu  $\hat{S}_T^W$ . Der Schätzer  $\hat{S}_T^{NW}$  ist positiv semi-definit und ein konsistenter Schätzer für  $S$ . Newey und West wählen mit  $[1 - (v/(m+1))]$  das sogenannte Bartlett Fenster. Dieses gewichtet die  $v = 1$  bis  $m$  Autokovarianzen mit linear abnehmenden Gewichten. Zur Verdeutlichung soll die Berechnung der  $v$ -ten Autokovarianzmatrix  $\hat{\Pi}_{v,T}$  in Matrixschreibweise dargestellt werden. Auch dieses Beispiel wird mit einer vom Modell vorgegebenen Momentenbedingung und einem  $(3 \times 1)$  Instrumentvektor berechnet.

$$\begin{aligned} \hat{\Pi}_{v,T} &= \frac{1}{T} \sum_{t=v+1}^T f(X_t, \hat{\theta}_T, z_t) f(X_{t-v}, \hat{\theta}_T, z_{t-v})' = \frac{1}{T} \sum_{t=v+1}^T (u_t z_t) (u_{t-v} z_{t-v})' \\ &= \frac{1}{T} \sum_{t=v+1}^T \begin{pmatrix} u_t \\ z_{1t} u_t \\ z_{2t} u_t \end{pmatrix} \begin{pmatrix} u_{t-v} & z_{1t-v} u_{t-v} & z_{2t-v} u_{t-v} \end{pmatrix} \\ &= \frac{1}{T} \begin{pmatrix} \sum u_t u_{t-v} & \sum z_{1t-v} u_t u_{t-v} & \sum z_{2t-v} u_t u_{t-v} \\ \sum z_{1t} u_t u_{t-v} & \sum z_{1t} z_{1t-v} u_t u_{t-v} & \sum z_{1t} z_{2t-v} u_t u_{t-v} \\ \sum z_{2t} u_t u_{t-v} & \sum z_{1t-v} z_{2t} u_t u_{t-v} & \sum z_{2t} z_{2t-v} u_t u_{t-v} \end{pmatrix} \end{aligned}$$

Zur Berechnung des Newey-West-Schätzers werden diese Autokovarianzmatrizen mit den entsprechenden Gewichten aufsummiert.

$S$  kann konsistent geschätzt werden, wenn  $m$ , die Bandbreite des Schätzers bzw. die Anzahl der berücksichtigten Lags, mit der Anzahl der Observationen steigt.<sup>42</sup> Für empirische Arbeiten ergibt sich das Problem, wie groß die Bandbreite des Schätzers, bei gegebenem Stichprobenumfang  $T$ , sein soll. Simulationsstudien haben gezeigt, daß sich die Güte von Parameter tests verbessert, wenn die optimale Bandbreite  $m$  gemäß eines von Newey und West dargestellten Algorithmus (automatische Bandbreitenwahl)

<sup>39</sup> Sie preferieren den Newey-West-Schätzer in Verbindung mit der automatischen Bandbreitenwahl und *pre-whitened* Residuen. Vgl. Andersen/Sørensen, 1996, S.349.

<sup>40</sup> Vgl. Davidson/MacKinnon, 1993, S.611.

<sup>41</sup> Vgl. Newey/West, 1994, S.633.

<sup>42</sup> Vgl. Hamilton, 1994, S.281.

bestimmt wird, der neben dem Stichprobenumfang  $T$  auch die Autokorrelationsstruktur der Daten berücksichtigt.<sup>43</sup> Mit größeren Autokovarianzen der Residuen  $\varepsilon_t$  und  $\varepsilon_{t-v}$  sowie zunehmender Bedeutung derselben auch für größere Lags  $v$ , wird die optimale Bandbreite des Newey-West-Schätzers steigen.<sup>44</sup>

*Pre-whitened* Residuen werden dann verwendet, wenn nicht ausgeschlossen werden kann, daß autoregressive Prozesse für die Autokorrelation in den Residuen verantwortlich sind. Der Newey-West-Schätzer für die Matrix  $S$  wird auf die transformierten Residuen angewandt. Die Berechnung mit Hilfe dieser Residuen, deren Autokorrelation im Vergleich zu den ursprünglichen  $\varepsilon_t$ 's geringer ist, gewährleistet, daß später berechnete Konfidenzintervalle für die Modellparameter  $\theta$  mit größerer Wahrscheinlichkeit den wahren Parameterwert beinhalten, als ohne diese Transformation.<sup>45</sup>

## 2.3 Testen im Rahmen der GMM

Im Anschluß an eine ökonometrische Schätzung werden aus zweierlei Gründen Tests durchgeführt. Zum einen werden die geschätzten Parameter daraufhin überprüft, ob Hypothesen bzgl. ihrer wahren Werte abgelehnt oder akzeptiert werden können. Zum anderen können Tests dazu dienen, die Gültigkeit der getroffenen Modellannahmen zu testen. Das Ensemble dieser Tests wird herangezogen, um eine zu favorisierende Spezifikation zu identifizieren.

Die GMM trifft keine Modellannahmen bzgl. der Residuen, deren Verletzung man testen könnte. Um Rückschlüsse bzgl. Fehlspezifizierungen zu ziehen, testet man daher auf die Gültigkeit überidentifizierender Restriktionen und auf Parameterstabilität.<sup>46</sup> Zudem existiert bei der GMM kein dem  $R^2$  äquivalentes Maß. Dessen Funktion wird z.T. ebenfalls vom Test auf überidentifizierende Restriktionen übernommen.

### 2.3.1 Parametertests

#### 2.3.1.1 Signifikanz der Parameter

Von großem Interesse ist bei den meisten Schätzungen, ob sich der Schätzer eines Parameters signifikant von Null oder einem anderen theoretisch vorgegebenen Wert unterscheidet. Im Rahmen von Kleinstquadratschätzungen dient der t-Test zur Überprüfung diesbezüglicher Hypothesen.

---

<sup>43</sup> Vgl. Newey/West, 1994, S.632f. Eine detailliertere Darstellung dieses Verfahrens befindet sich in Anhang A.

<sup>44</sup> Optimal bedeutet in diesem Zusammenhang, daß der Schätzer ein Kriterium aus Verzerrung und Varianz minimiert.

<sup>45</sup> Vgl. Andrews/Monahan, 1992, Newey/West, 1994 und Andersen/Sørensen, 1996. Zu den technischen Details der Transformation: Vgl. Andrews/Monahan, 1992, S.954f und Anhang B.

<sup>46</sup> Vgl. Davidson/MacKinnon, 1993, S.616.

Die Parameter einer GMM-Schätzung sind approximativ normalverteilt,  $\sqrt{T}(\hat{\theta}_T - \theta_0) \sim N(0, V)$ .<sup>47</sup> Hieraus folgt, daß  $\sqrt{T}(\hat{\theta}_T - \theta_0)/\sqrt{V} \sim N(0,1)$ , d.h. standardnormalverteilt und der Schätzer  $\hat{\theta}_T \sim N(\theta_0, V/T)$  verteilt ist. Der Mittelwert der Verteilung ist der wahre Parameter  $\theta_0$  und die Varianz berechnet sich mit Hilfe von  $Cov(S^{-1}) = (\Gamma' S^{-1} \Gamma)^{-1} = V$ . Die Matrix  $S$  ist ein im Abschnitt 2.2 beschriebener HAC-Schätzer für die Varianz-Kovarianzmatrix der Orthogonalitätsbedingungen.<sup>48</sup>

Es lassen sich für jeden einzelnen Parameter im Parametervektor  $\theta$  t-Werte berechnen, indem man den Schätzwert durch die Standardabweichung dividiert. Diese erhält man aus der Wurzel des entsprechenden Diagonalelements der Matrix  $V/T$ .<sup>49</sup> Die t-Werte werden dann mit den kritischen Werten der Standard-Normalverteilung verglichen, um auf Signifikanz der Parameter zu testen.

### 2.3.1.2 Restriktionen über mehrere Parameter

Zur Überprüfung mehrerer oder nichtlinearer Parameterrestriktionen bieten sich Tests vom Lagrange Multiplier (LM), Likelihood Ratio (LR) oder Wald Typ an, die ursprünglich zur Anwendung bei ML-Schätzungen entwickelt wurden.<sup>50</sup> Auf den entsprechenden LM-Test soll in dieser Arbeit nicht weiter eingegangen werden, da dessen Anwendung im Rahmen einer GMM-Schätzung zu aufwendig ist.

Mit Hilfe eines Wald-Tests kann die Gültigkeit einer allgemeinen Parameterrestriktion getestet werden.<sup>51</sup> Ein Test vom Wald-Typ prüft, wie stark unrestringierte Schätzer die Parameterrestriktionen verletzen.<sup>52</sup> Diese Restriktion sei gegeben durch den  $(s \times 1)$  Vektor  $a(\theta) = 0$ . Bei Gültigkeit der Restriktion sollte diese auch für den unrestringiert geschätzten Parametervektor  $\hat{\theta}_T$  gültig sein. Die Wald-Test Prüfgröße lautet:

$$W = T a(\hat{\theta}_T)' \hat{V}_{Ta}^{-1} a(\hat{\theta}_T)$$

$$\text{mit } \hat{V}_{Ta} = \left( \frac{\partial a(\hat{\theta}_T)}{\partial \theta'} \right) \left( \hat{\Gamma}_T' \hat{S}_T^{-1} \hat{\Gamma}_T \right)^{-1} \left( \frac{\partial a(\hat{\theta}_T)}{\partial \theta'} \right)'$$

$\hat{V}_{Ta}$  ist die geschätzte Varianz-Kovarianzmatrix des Restriktionsvektors  $a(\hat{\theta}_T) = 0$ .

$(\partial a(\hat{\theta}_T)/\partial \theta') = \Lambda_T$  ist die  $(q \times s)$  Matrix der partiellen Ableitungen der Restriktionen nach den Parametern, bewertet zu den unrestringiert geschätzten Parameterwerten. Um

<sup>47</sup> Vgl. Abschnitt 2.1.4.

<sup>48</sup> Vgl. Hamilton, 1994, S.415.

<sup>49</sup> Vgl. Hall, 1993, S.402.

<sup>50</sup> Alle diese Tests sind nur für große Stichproben gültig, vgl. Kmenta, 1996, S.491.

<sup>51</sup> Vgl. Newey/West, 1987b, S.780.

<sup>52</sup> Vgl. Kmenta, 1986, S.492.

die Restriktion mit Hilfe des Wald-Tests zu überprüfen, muß also lediglich die unrestringierte Form geschätzt werden. Die Nullhypothese, Gültigkeit der Restriktion, wird abgelehnt, wenn der berechnete Wert der  $\chi^2$  (s) verteilten Prüfgröße über dem der theoretischen Verteilung liegt.

Auch ein LR-Test kann zum Testen auf Parameterrestriktionen benutzt werden. Gilt die Nullhypothese der Restriktion, d.h.  $a(\theta) = 0$ , dann sollten die im Zuge der GMM-Schätzung zu minimierenden Zielfunktionen  $J_T(\hat{\theta}_T)$  und  $J_T(\hat{\theta}_T^r)$ , mit  $\theta_T^r$  als restringiertem Parametervektor, ähnlich niedrige Werte annehmen. Dies kann man mit Hilfe des LR Tests testen, dessen Prüfgröße  $\chi^2$ (s) verteilt ist und wie folgt aussieht:

$$LR = T(J_T(\hat{\theta}_T^r) - J_T(\hat{\theta}_T)). \quad (6)$$

Um diesen Test durchführen zu können, müssen sowohl die unrestringierte als auch die restringierte Schätzung durchgeführt werden. Zu beachten ist, daß bei der Berechnung beider Zielfunktionen derselbe Schätzer für die Matrix  $S$  benutzt wird, beide Schätzungen folglich mit derselben Gewichtungsmatrix arbeiten. Diese Anforderung verursacht im Rahmen einer iterativen GMM-Schätzung Schwierigkeiten. Wird die Gewichtungsmatrix, die z.B. aus der letzten Iteration der unrestringierten Schätzung resultiert, zur Schätzung des restringierten Prozesses benutzt, ist nicht zu erwarten, daß die resultierenden Parameter diejenigen sind, die man mit einer iterativen Schätzstrategie erhalten würde. Ogaki schlägt vor, die resultierende Gewichtungsmatrix des einen Prozesses als erste Gewichtungsmatrix im zweiten zu verwenden. In weiteren Iterationen kommt dann die jeweils neu geschätzte Gewichtungsmatrix zur Anwendung, bis das Ergebnis konvergiert.<sup>53</sup> Dieser Vorschlag ist allerdings nicht geeignet, die Anforderung gleicher Gewichtungsmatrizen bei der Berechnung der Zielfunktionen  $J_T(\hat{\theta}_T)$  und  $J_T(\hat{\theta}_T^r)$  zu erfüllen, da sich die Gewichtungsmatrizen der jeweils letzten Schätzungen unterscheiden.

Da bei Gültigkeit der Nullhypothese,  $a(\theta) = 0$ , die Gewichtungsmatrix aus der unrestringierten Schätzung auch ein konsistenter Schätzer für die optimale Gewichtungsmatrix der restringierten Schätzung ist, wird in dieser Arbeit beim LR Test wie folgt vorgegangen: Das unrestringierte Modell wird iterativ geschätzt. Die resultierende Gewichtungsmatrix wird dann als Gewichtungsmatrix in der restringierten Schätzung eingesetzt. Aus dieser Schätzung resultiert das  $J_T(\hat{\theta}_T^r)$  in Gleichung (6).<sup>54</sup>

Alle drei Tests, Wald-, LR- und LM-Test, sind asymptotisch äquivalent. Lediglich bei kleinen Stichproben unterscheidet sich die Macht der Tests. Man kann folglich auf

---

<sup>53</sup> Vgl. Ogaki, 1993b, S.14.

Grund der Verfügbarkeit der notwendigen Komponenten entscheiden, welchen der drei Tests man anwendet.

### 2.3.2 Überidentifizierende Restriktionen

Ist die Anzahl der Orthogonalitätsbedingungen größer als die Anzahl der zu schätzenden Parameter ( $r > q$ ), dann stehen mehr Restriktionen zur Verfügung als man zur Schätzung des Parametervektors  $\theta$  benötigt. Wenn das Modell korrekt spezifiziert ist, würde man von einem guten Schätzer dennoch erwarten, daß alle, bei der Minimierung von (2) verwendeten, Orthogonalitätsbedingungen erfüllt werden, d.h.  $g(\hat{\theta}_T, x_t, z_t)$  nahe bei Null liegt.<sup>55</sup> Hansen schlägt folgenden Test vor, um dies zu überprüfen.<sup>56</sup> Ausgehend von der Größe  $\sqrt{T}g(\hat{\theta}_T, x_t, z_t)$ , die normalverteilt ist, konstruiert er die Prüfgröße

$$\begin{aligned} & [\sqrt{T}g(\hat{\theta}_T, x_t, z_t)]' \hat{S}_T^{-1} [\sqrt{T}g(\hat{\theta}_T, x_t, z_t)] \\ &= Tg(\hat{\theta}_T, x_t, z_t)' \hat{S}_T^{-1} g(\hat{\theta}_T, x_t, z_t) \\ &= TJ_T(\hat{\theta}_T). \end{aligned}$$

Von dieser Prüfgröße, Stichprobenumfang multipliziert mit dem Wert der Zielfunktion, weiß man, daß sie  $\chi^2$  verteilt ist und  $(r - q)$  Freiheitsgrade besitzt.<sup>57</sup> Die Nullhypothese besagt, daß  $g(\theta_0, x_t, z_t)$ , der Vektor der Orthogonalitätsbedingungen in der Grundgesamtheit, gleich Null ist. Übersteigt die berechnete Prüfgröße den entsprechenden kritischen Wert aus der  $\chi^2$ -Verteilung, dann muß diese Nullhypothese abgelehnt werden. Wird  $H_0$  abgelehnt, bedeutet dies, daß das Modell mißspezifiziert ist. Newey warnt allerdings davor, diesen Test als einen generellen Test auf Mißspezifikation zu betrachten, da sie nachweist, daß der Test nicht auf jede Art der Modell-Mißspezifikation sensibel reagiert.<sup>58</sup>

Chan et al. benutzen diesen Test auch als ein Maß für die Güte der Schätzung, mit dessen Hilfe die Güte von Schätzungen verglichen werden kann. Je größer der p-value zum  $\chi^2$ -Test, desto höher ist der Erklärungswert des Modells.<sup>59</sup>

Abgesehen von Neweys obigem Einwand, hängt die Güte dieses Tests auch vom Verhältnis zwischen der Anzahl der Observationen und der Anzahl der Orthogonalitätsbedingungen ab. Andersen und Sørensen zeigen in ihrer Simulationsstudie, daß der Ein-

---

<sup>54</sup> Diese Vorgehensweise wird von Torben G. Andersen vorgeschlagen. e-mail vom 7.März 1998.

<sup>55</sup> Vgl. Hamilton, 1994, S.415.

<sup>56</sup> Vgl. Hansen, 1982, S.1049-1050.

<sup>57</sup> Eine Prüfgröße der Form  $X = \sum_{i=1}^n g_i^2$  ist  $\chi^2$  verteilt mit n Freiheitsgraden. Genau q Linearkombinationen der Elemente in g sind gleich 0. Das folgt aus der Minimierungsbedingung der GMM. Die Prüfgröße besitzt demzufolge  $(r - q)$  Freiheitsgrade. Vgl. Härtter, 1987, S.148.

<sup>58</sup> Vgl. Newey, 1985, S.248.

satz zuvieler Orthogonalitätsbedingungen dazu führt, daß falsche Nullhypothesen zu selten verworfen werden. Umgekehrt gilt, daß bei Verwendung einer zu geringen Zahl von Instrumenten eine richtige Nullhypothese zu häufig verworfen wird. Dies wird von den Ergebnissen Tauchens bestätigt, der bei einer sehr kleinen Stichprobe zu dem Ergebnis kommt, daß der Test eher dazu neigt, eine Modellspezifikation zu akzeptieren. Ein ablehnendes Testurteil sei in diesem Fall ein sehr starkes Indiz für einen Spezifikationsfehler.<sup>60</sup>

Für die praktische Arbeit lassen sich aus diesen Ergebnissen allerdings nur qualitative Schlüsse ziehen. Eine Faustregel zur Bestimmung der optimalen Anzahl der zu benutzenden Instrumente kann nicht abgeleitet werden.

### 2.3.3 Strukturbruch

Wegen der angesprochenen Schwächen des  $\chi^2$ -Tests empfiehlt es sich, weitere Tests heranzuziehen, um die Güte der Schätzung und die Gültigkeit der Spezifikation zu beurteilen. Diesem Zweck kann auch ein Strukturbruchtest dienen. Wird ein Strukturbruch diagnostiziert, ohne daß es für diesen eine plausible ökonomische Erklärung gibt, könnte eine Fehlspezifikation der Grund dafür sein, daß das Modell nicht in der Lage ist, den Prozeß der Daten über den gesamten Beobachtungszeitraum gut zu erklären.<sup>61</sup>

Chan et al. testen auf einen Strukturbruch, indem sie für jeden der zu schätzenden Parameter eine zusätzliche Dummy Variable einführen und testen, ob die Nullhypothese der Insignifikanz dieser Parameter verworfen werden kann. Im Kern wird somit ein Test auf eine Parameterrestriktion durchgeführt, der schon besprochen wurde.<sup>62</sup>

Hamilton beschreibt den folgenden Strukturbruchtest.<sup>63</sup> Die gesamte Stichprobe wird in zwei Teilstichproben getrennt und folgende Dummy-Variable wird eingeführt: ( $T_0$  ist die erste Beobachtungswert der zweiten Teilstichprobe)

$$d_t = \begin{cases} 1 & \text{für } t \leq T_0 \\ 0 & \text{für } t > T_0 \end{cases} .$$

Die Orthogonalitätsbedingung  $f(\theta, x_t, z_t)$  kann dann mit dieser Dummy-Variablen erweitert werden. Es entsteht ein  $(2r \times 1)$  Vektor mit zwei getrennten Sets von insgesamt  $2r$  Orthogonalitätsbedingungen.

---

<sup>59</sup> Vgl. Chan et al., 1992, S.1217 und Andersen/Sørensen, 1996, S.344-347.

<sup>60</sup> Vgl. Tauchen, 1986, S.412 und Andersen/Sørensen, 1996, S.344ff.

<sup>61</sup> Vgl. Ghysels/Hall, 1990, S.122f.

<sup>62</sup> Vgl. Chan et al., 1992, S.1220-1222.

<sup>63</sup> Vgl. Hamilton, 1994, S.424.

$$f^*(\theta, x_t, z_t, d_t) \equiv \begin{bmatrix} f(\theta, x_t, z_t) \cdot d_t \\ f(\theta, x_t, z_t) \cdot (1 - d_t) \end{bmatrix}.^{64}$$

Mit diesen neu konstruierten Orthogonalitätsbedingungen erhält man einen neuen Schätzer für  $\theta - \hat{\theta}_{T,T_0}$ . Dieser Schätzer sollte, soweit kein Strukturbruch vorliegt, beide Sets von Orthogonalitätsbedingungen erfüllen. Unterscheiden sich die wahren Parameter der beiden Teilperioden ( $\theta_1$  und  $\theta_2$ ), dann ist zu erwarten, daß der Test bzgl. der überidentifizierenden Restriktionen die Nullhypothese

$$g^*_T(\theta) = (1/T) \sum_{t=1}^T f^*(\theta_1 = \theta_2, x_t, z_t, d_t) = 0 \quad (7)$$

ablehnen wird.<sup>65</sup> Ein einziger Schätzer wird es nicht bewerkstelligen können, alle Orthogonalitätsbedingungen für im Kern zwei unterschiedliche Strukturen zu erfüllen. Andererseits sollte, wenn  $\theta_1 = \theta_2$  gilt, die Nullhypothese (7) nicht abgelehnt werden. Dieser Test verliert seine Aussagekraft wenn die Teilstichproben sehr unterschiedlich lang sind. Ist dies der Fall, dann wird die Parameterschätzung und der Wert der Funktion  $J_T$  zu sehr von dem längeren Zeitraum dominiert, und selbst bei sehr verschiedenen Parametern  $\theta_1$  und  $\theta_2$  wird es immer unwahrscheinlicher, daß dieser Test einen Strukturbruch erkennt. Ein weiteres Problem dieses Tests besteht darin, daß er die Anzahl der Orthogonalitätsbedingungen konstruktionsbedingt verdoppelt. Wie bereits erwähnt, reagiert die GMM Schätzmethode u.U. sehr sensibel auf die Auswahl und die Anzahl der Instrumente. Die Ergebnisse der Tests sind folglich mit der gebührenden Vorsicht zu interpretieren. Dies gilt zusätzlich, da auch dieser Test auf der Annahme der Stationarität der Zeitreihen beruht, damit der Zentrale Grenzwertsatz anwendbar ist.<sup>66</sup> Im folgenden sei dieser Test D-Strukturbruchttest genannt.

Andrews und Fair beschreiben Wald-, LM- und LR- Strukturbruchttests. Im Rahmen dieser Arbeit wird auf diese nicht näher eingegangen, da sie entweder für eine GMM-Schätzung nur unter sehr restriktiven Bedingungen gültig, bzw. einer Berechnung im Rahmen von RATS nur schwer zugänglich sind.<sup>67</sup>

Strukturbruchttests werden in der empirischen Literatur im Rahmen von GMM-Schätzungen, trotz ihrer Bedeutung für die Modellbeurteilung, meist vernachlässigt.<sup>68</sup>

<sup>64</sup> Dieser Vektor besitzt nun  $2r$  Orthogonalitätsbedingungen und  $\theta$  ist damit auf jeden Fall überidentifiziert.

<sup>65</sup> Da ein Test auf überidentifizierende Restriktionen durchgeführt wird, sind die entsprechenden Kritikpunkte auch an dieser Stelle zu beachten.

<sup>66</sup> Dies gilt, wenn Zweifel bzgl. der Stationarität der Zeitreihen existieren. So etwa im behandelten Anwendungsbeispiel. Vgl. Abschnitt 4.1.2.

<sup>67</sup> Vgl. Andrews/Fair, 1988, S.624.

<sup>68</sup> Ghysels und Hall testen, systematisch auf Strukturbrüche, kommen aber mit verschiedenen Tests zu unterschiedlichen Ergebnissen. Vgl. Ghysels/Hall, 1990, S.134 und speziell Fußnote 4.

## 2.4 Zwischenfazit zur GMM

An dieser Stelle läßt sich nochmals der große Vorteil der GMM darlegen und aufzeigen, warum man akzeptiert, einen Parameter zwar konsistent, in großen Stichproben aber ineffizient, zu schätzen.<sup>69</sup> Gehen wir z.B. von der OLS aus, die unter strengen Annahmen BLU-Schätzer liefert. Damit die resultierenden Schätzer diese Eigenschaften haben, müssen die Gauss-Markov Bedingungen erfüllt sein, die unter anderem fordern, daß die Residuen homoskedastisch und nicht autokorreliert sind. Sind diese Voraussetzungen nicht erfüllt, dann verlieren die OLS-Schätzer ihre BLU-Eigenschaften und, was noch schwerer wiegt, Rückschlüsse von den Schätzern auf die Parameter der Grundgesamtheit werden unmöglich, da die Parametervarianzen verzerrt geschätzt werden.

Ist es unmöglich Parameter mit BLU-Eigenschaften zu erhalten, und ist man außerdem bereit, den Großstichprobeneigenschaften der Schätzer (Konsistenz und asymptotische Effizienz) den größten Wert beizumessen, dann sind die ML-Methode und die GMM eine naheliegende Wahl. Die ML wird vor allem dann ausgewählt, wenn die Verteilung der Zufallsvariablen bekannt ist, da die ML-Schätzer dann sowohl konsistent als auch effizient sind. Allerdings verlieren die ML-Schätzer diese Eigenschaften, wenn die getroffenen Verteilungsannahmen nicht erfüllt werden. In vielen Fällen, so z.B. im nachfolgenden Anwendungsbeispiel, ist die Kenntnis der Verteilung nicht selbstverständlich. Die GMM bietet die Möglichkeit, die Modellparameter auch ohne die Kenntnis der Verteilung konsistent zu schätzen.

Unter welchen Voraussetzungen dieses Ergebnis abgeleitet werden konnte, soll nochmals kurz rekapituliert werden.

Der wahre Parametervektor  $\theta_0$  muß der einzige Vektor sein, für den  $J(\theta) = 0$  gilt. Es darf keinen zweiten Parametervektor geben, der eben diese Bedingung erfüllt.<sup>70</sup>

Die für Anwendungen kritischste Annahme ist diejenige der Stationarität der Variablen, d.h. des Regressors, der Regressanden und der Instrumentvariablen. Eine Variable ist dann stationär, wenn ihr erwarteter Mittelwert und ihre erwartete Varianz über die Zeit konstant sind, und die Kovarianz  $E(x_t x_{t-\tau}) =$  konstant ist für alle  $t$  und nur mit  $\tau$  variiert.<sup>71</sup> Im Zusammenhang mit der GMM ist es von Bedeutung, daß diese Bedingungen nur für die unbedingten Momente gelten. Letztere Einschränkung läßt die E-

---

<sup>69</sup> Vgl. Johnston/DiNardo, 1997, S.333.

<sup>70</sup> Vgl. Ogaki, 1993a, S.464 und S.476 für einen Fall bei dem diese Voraussetzung verletzt wird, da das Modell so formuliert ist, daß eine bestimmte Parameterkonstellation die Orthogonalitätsbedingung automatisch und für alle denkbaren Variablenwerte erfüllt.

<sup>71</sup> Vgl. Hamilton, 1994, S.45f.

xistenz von bedingter Heteroskedastizität und serieller Autokorrelation zu. Dies sind häufig beobachtbare Charakteristika ökonomischer Datenreihen.<sup>72</sup>

Hall stellt fest,<sup>73</sup> daß die Annahme der Stationarität im Rahmen der GMM lediglich notwendig ist, um die Ableitung der Verteilung der Parameter und den Beweis der Parameterkonsistenz (mittels Gesetz der großen Zahlen und Zentralem Grenzwertsatz) zu erleichtern. Die Annahme der Stationarität sei hierzu allerdings nicht zwingend notwendig, schreibt Hall, bezugnehmend auf Gallant und White, die an Datenreihen weniger strenge Voraussetzungen stellen, um das Gesetz der großen Zahlen und den Zentralen Grenzwertsatz anzuwenden. Dies sind Anforderungen bezüglich der Unabhängigkeit verschiedener Stichproben aus derselben Datenreihe und die Forderung, daß ein Prozeß nicht zu sehr von weit zurückliegenden Daten abhängt.<sup>74</sup> Dies ist ein Grund dafür, daß in empirischen Arbeiten, welche die GMM anwenden, die Annahme der Stationarität der Datenreihen eher flüchtig, wenn überhaupt, geprüft wird.

### 3 Empirischer Teil - Schätzung des Zinsprozesses mittels GMM

Die Anwendung der GMM wird anhand eines Beispiels aus der empirischen Kapitalmarktforschung verdeutlicht. Es werden die Parameter eines stochastischen Zinsprozesses geschätzt. Die Modellierung des kurzfristigen und risikolosen Zinssatzes<sup>75</sup> als stochastischer Prozess hat speziell im Rahmen von Zinsstrukturmodellen zentrale Bedeutung.<sup>76</sup> Dem Beispiel von Chan, Karoly, Longstaff und Sanders (CKLS) folgend, dient die stochastische Differentialgleichung

$$di = (\alpha + \beta i)dt + \psi i^\gamma dz \quad (8)$$

als Ausgangspunkt der Schätzung.<sup>77</sup> Unterschiedliche Modelle unterscheiden sich in den Restriktionen, die sie den Prozessparametern  $\alpha$ ,  $\beta$ ,  $\gamma$  und  $\psi$  auferlegen.

Die Schätzung der Spezifikation (8) mittels GMM ist vorteilhaft, weil keine Verteilungsannahme bzgl. der Zinsänderungen getroffen werden muß. Führt man eine ML-Schätzung durch, müßte a priori festgelegt werden, ob  $di$  z.B. normal- oder t-verteilt ist. Diese Eigenschaft der GMM ist im vorliegenden Anwendungsbeispiel gerade deshalb von Vorteil, weil verschiedene Werte des Parameters  $\gamma$  unterschiedliche Verteilungen

---

<sup>72</sup> Vgl. Ogaki, 1993a, S.463. So wird die Varianz der Zinsänderung im später behandelten Beispiel z.B. als eine Funktion des Zinsniveaus modelliert, sie ist folglich bedingt heteroskedastisch.

<sup>73</sup> Vgl. Hall, 1993, S.413.

<sup>74</sup> Davidson/MacKinnon, 1993, S.132.

<sup>75</sup> In dieser Arbeit modelliert mit dem 1-Monats-Geldmarktsatz.

<sup>76</sup> Gemeint sind hiermit sogenannte Gleichgewichtsmodelle der Zinsstruktur. Im Rahmen dieser Arbeit kann nicht näher auf Zinsstrukturmodelle und die Einbindung des Zinsprozesses in diese eingegangen werden. Es wird auf Hull, 1991, Kap.15, Rebonato, 1996, Kap.9 und Campbell et al., 1997, Kap.11 verwiesen.

<sup>77</sup> Vgl. Chan et al., 1992, S. 1210 und Baxter/Rennie, 1996, S.54ff zu stochastischen Differentialgleichungen im allgemeinen.

für die Zinsänderungen implizieren. Gälte  $\gamma = 0$ , so wären Zinsänderungen normalverteilt. Besäße das Cox / Ingersoll / Ross (CIR) Modell Gültigkeit, welches  $\gamma = 0,5$  setzt, so folgten die Zinsänderungen einer  $\chi^2$ -Verteilung.<sup>78</sup> Aus dem Brennan / Schwartz Modell folgen mit der Restriktion  $\gamma = 1$  log-normalverteilte Zinsänderungen.<sup>79</sup>

Zwei Implikationen dieser Spezifikation für den Zinsprozeß sollen noch kurz erläutert werden. Zum einen erlaubt sie, den sogenannten *mean-reversion* Effekt zu modellieren. Der Zins nähert sich mit der Rate  $\beta$  dem *mean-reversion* Niveau,  $\alpha / \beta$ , an. Die zweite Eigenschaft ist die zinsniveauabhängige Volatilität der Zinsänderung  $\psi i^\gamma$ . Gilt  $\gamma > 0$ , dann steigt die Volatilität mit einem höheren Zinsniveau. Sowohl der *mean-reversion* Effekt, als auch eine höhere Volatilität bei hohem Zinsniveau (LEVEL-Effekt) sind empirisch beobachtbare Eigenschaften des Zinsprozesses.<sup>80</sup>

### 3.1 Formale Darstellung der Schätzproblematik

Die Darstellung der GMM im theoretischen Teil der Arbeit basierte auf der Annahme, daß das theoretische Modell eine Momentbedingung vorgibt, und weitere Orthogonalitätsbedingungen durch die Interaktion mit Instrumenten entstehen. Bei der Schätzung des Prozesses des kurzfristigen, risikolosen Zinssatzes gemäß Gleichung (8) stehen Aussagen zum ersten und zum zweiten Moment der Zinsänderung zur Verfügung. Die Drift sei  $(\alpha + \beta i)$  und die Standardabweichung  $\psi i^\gamma$ .

Das in kontinuierlicher Zeit beschriebene Modell muß nun durch diskrete Daten approximiert werden. Dies geschieht mit der folgenden Spezifikation:

$$\Delta i_t = \alpha + \beta i_{t-1} + v_t$$

mit  $E(v_t | \Omega_{t-1}) = 0$ ,  $E(v_t^2 | \Omega_{t-1}) \equiv \sigma_t^2 = \psi^2 i_{t-1}^{2\gamma}$  und  $\Omega_{t-1}$  als der zum Zeitpunkt  $t-1$  zur Verfügung stehenden Informationsmenge.<sup>81</sup>

Die aus dem Modell abgeleiteten Momentbedingungen ergeben sich aus den Erwartungen über die Zinsänderung und die Varianz:

$$u_t = h(x_t, \theta) = h \begin{pmatrix} \Delta i_t - \alpha - \beta i_{t-1} \\ (\Delta i_t - \alpha - \beta i_{t-1})^2 - \psi^2 i_{t-1}^{2\gamma} \end{pmatrix} = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

mit  $u_t$  als dem  $(2 \times 1)$  Vektor der Residuen und dem Parametervektor  $\theta = (\alpha \beta \psi^2 \gamma)$ . Der Vektor  $x_t$  stehe auch weiterhin für den Vektor aller exogenen und endogenen Variablen zum Zeitpunkt  $t$ . Da es vier Parameter zu schätzen gilt, benötigt man mindestens eben-

<sup>78</sup> Vgl. Chan et al., 1992, S.1214.

<sup>79</sup> Vgl. Rebonato, 1996, S.301.

<sup>80</sup> Vgl. Rebonato, 1996, S.175 und die Darstellung der Datenreihe im Anhang C.

<sup>81</sup> Vgl. Brenner et al., 1996, S.87 und z.B. Andersen/Lund, 1997, S.345 zu den Problemen die mit diesem Schritt verbunden sind.

so viele Orthogonalitätsbedingungen. Diese Arbeit folgt der gängigen Praxis und benutzt verzögerte Variablen aus der Informationsmenge  $\Omega_{t-1}$  als Instrumente.<sup>82</sup> Im folgenden wird mit dem  $(4 \times 1)$  Instrumentvektor  $z_t = (1 \ i_{t-1} \ i_{t-2} \ i_{t-3})'$  gearbeitet. Der  $((2 * 4) \times 1) = (8 \times 1)$  Vektor der Orthogonalitätsbedingungen entsteht dann aus dem Kronecker Produkt der Momentbedingungen und des Instrumentvektors

$$f(x_t, \theta, z_t) = h(x_t, \theta) \otimes z_t \\ = (u_{1,t} \ u_{1,t}i_{t-1} \ u_{1,t}i_{t-2} \ u_{1,t}i_{t-3} \ u_{2,t} \ u_{2,t}i_{t-1} \ u_{2,t}i_{t-2} \ u_{2,t}i_{t-3})'$$

Es bleiben folglich vier überidentifizierende Restriktionen. Die GMM-Schätzung verläuft ansonsten gemäß der Darstellung im theoretischen Teil für eine Momentbedingung. Da dieses Modell die Volatilität vom Zinsniveau abhängig macht, sei es im folgenden als LEVEL-Modell bezeichnet.

### 3.2 Ergebnisse zum LEVEL Modell

Im folgenden werden die Ergebnisse der GMM-Schätzung zum LEVEL-Modell dargestellt. Es kommt die in Abschnitt 2.1.2 beschriebene, iterative Schätzstrategie zur Anwendung.<sup>83</sup> Der Instrumentvektor  $z_t$  sei der in Abschnitt 3.1 beschriebene  $(4 \times 1)$  Vektor, es wird folglich mit acht Orthogonalitätsbedingungen geschätzt. Da die Residuen  $\varepsilon_t$  autokorreliert sind, wird die Varianz-Kovarianzmatrix  $S$  mit Hilfe des Newey-West-Schätzers bestimmt.<sup>84</sup> Dabei wird die Bandbreite des Schätzers mit Hilfe des in Abschnitt 2.2.2 vorgestellten Verfahrens zur Ermittlung der Schätzerbandweite ermittelt.<sup>85</sup> Da die AKF und PAKF der  $\varepsilon_t$ s weithin auf das Vorhandensein von autoregressiven Prozessen in den Residuen der Orthogonalitätsbedingungen hindeuten, soll im folgenden der Newey-West-Schätzer auf die *pre-whitened* Residuen angewandt werden.

Es wurde geprüft, ob die Zinsdaten (Wochenschlußkurse vom 2.1.1970 bis zum 20.2.1998) der Stationaritätsanforderung der GMM genügen. Entsprechend durchgeführte Augmented-Dickey-Fuller Tests können die Nullhypothese der Nichtstationarität nicht verwerfen. Dies entspricht den Ergebnissen anderer Autoren, die die Stationa-

<sup>82</sup> Vgl. z.B. Hansen/Singleton, 1982 und Chan et al., 1992.

<sup>83</sup> I.d.R. wurden nach spätestens acht Iterationen Schätzer gefunden, die sich bis auf die dritte Nachkommastelle nicht mehr verändert haben.

<sup>84</sup> Es wurden LM Tests durchgeführt, um auf Autokorrelation zu prüfen. Die Ergebnisse zeigen deutlich das Vorhandensein von Autokorrelation an.

<sup>85</sup> Die vom *automatic-bandwidth-choice* Verfahren berechnete optimale Bandbreite des Tests lag, bei vollem Stichprobenumfang um 20.

rität von Zinsreihen überprüfen.<sup>86</sup> Gerechtfertigt wird die Anwendung der GMM mit der Argumentation, daß es genüge, wenn sich die Zeitreihen nicht explosiv verhalten.<sup>87</sup> Wie Tabelle 1 mit den Ergebnissen zu entnehmen ist, ergeben sich bei der Schätzung des Grundmodells durchweg ökonomisch sinnvolle Ergebnisse. Das *mean-reversion* Niveau, das sich aus  $\alpha / \beta$  berechnen läßt, beträgt 5,42%. Der *mean-reversion* Parameter  $\beta$  ist negativ und signifikant.<sup>88</sup> Er gibt an, daß innerhalb einer Woche lediglich 0,5% der Differenz zwischen *mean-reversion* Niveau und  $r_{t-1}$  „aufgeholt“ wird. Der Parameter  $\gamma$ , der die Sensitivität der Volatilität bzgl. des Zinsniveaus beschreibt, wird mit 0,82 als signifikant von 0 verschieden geschätzt. Der Volatilitätsfaktor  $\psi^2$  bleibt insignifikant.<sup>89</sup>

**Tabelle 1:** Schätzergebnisse aus LEVEL und CIR Spezifikation. Freiheitsgrade und p-values in Klammern.  $R^2$  gibt den Anteil der Varianz der Zinsvolatilität an, der von der geschätzten, bedingten Volatilität erklärt wird und ist demnach ein Maß für die Prognosegüte. D75, D7583 und DSPEZ sind die Prüfgrößen für die D-Strukturbruchtests mit den Stichprobenumfängen 70-74,75-98 sowie 70-74,75-83 und 1/75-7/86,8/86-2/98.

	LEVEL	CIR	CIR-Dummy75
$\alpha$	0,03175 (0,02926)	0,02975 (0,03758)	0,03273 (0,03484)
$\beta$	-0,00586 (0,02020)	-0,00559 (0,02392)	-0,00609 (0,02963)
$\gamma$	0,82063 (0,00283)	0,5	0,5
$\psi^2$	0,00313 (0,34556)	0,00962 (0,00001)	0,04942 (0,00037)
$\psi^2-D$	-	-	-0,04532 (0,00369)
$\chi^2$	(4)=5,66704 (0,22543)	(5)=6,33511 (0,27496)	(4)=5,829717 (0,21223)
$R^2$	0,03592	0,01766	0,10038
D75: $\chi^2$	(12)=13,95 (0,30385)	(13)=14,14632 (0,363621)	(12)=12,60454 (0,39843)
D7583: $\chi^2$	(12)=15,50 (0,21529)	(13)=15,48965 (0,27779)	(12)=13,62007 (0,32562)
DSPEZ: $\chi^2$	(12)=15,73725 (0,20357)	(13)=15,52391 (0,27580)	(12)=14,42432 (0,27444)

Der  $\chi^2$ -Test auf überidentifizierende Restriktionen kann bei einem Signifikanzniveau von 0,22 die Nullhypothese, daß auch die überidentifizierenden Restriktionen erfüllt

<sup>86</sup> Vgl. Andersen/Lund, 1997, S.349 und Aftab-Sahalia, 1996, S.399.

<sup>87</sup> Vgl. Andersen/Lund, 1997, S.349. Auch in dieser Arbeit soll nicht näher auf dieses Problem eingegangen werden und die Tatsache, daß sich auch die Zeitreihe der deutschen 1-Monats-Geldmarktsätze (Anhang C) nicht explosiv verhält soll als Rechtfertigung für die Anwendung der GMM genügen.

<sup>88</sup> Hier und im folgenden impliziert „signifikant“ immer eine Irrtumswahrscheinlichkeit von 5%, es sei denn, es wird explizit auf ein anderes Signifikanzniveau hingewiesen. Ein signifikanter *mean-reversion* Effekt ist ein weiteres Indiz für eine nicht-explosive Zeitreihe.

<sup>89</sup> Ist die Volatilität vom Zinsniveau abhängig, läßt sich  $\psi$  nicht als alleinstehender Parameter interpretieren.

sind, nicht ablehnen. Ebenso wie der DSPEZ-Test<sup>90</sup> liefert dieses Ergebnis somit keinen Hinweis auf eine Fehlspezifikation. Allerdings deutet die Tatsache, daß lediglich 3,6% der Zinsvolatilität von diesem Modell prognostiziert werden können, darauf hin, daß diese Spezifikation den Zinsprozeß nur unzureichend erklären kann.<sup>91</sup> Weder der D75- bzw. der D7583- Strukturbruchtest, noch zusätzlich eingeführte Dummy-Variablen<sup>92</sup> für  $\gamma$  und  $\psi^2$  erweisen sich als signifikant. Die Nullhypothese dieser Tests, das Nichtvorhandensein eines Strukturbruches zum Jahreswechsel 1974/75, kann folglich nicht abgelehnt werden.

Es soll geprüft werden, ob die Zinsen einem Random Walk folgen. Dieser ist eine Prozeßart, die häufig auf Finanzmarktdatenreihen zutrifft. Gälten die Restriktionen  $\alpha = \beta = 0$ , so folgte der kurzfristige Zinssatz einem Random Walk ohne Drift. Eine solche Diagnose hätte auch zur Folge, daß die Anwendbarkeit der GMM als Schätzmethode für das vorliegende Problem weiter in Frage gestellt würde, da ein Random Walk das klassische Beispiel einer nichtstationären Zeitreihe ist. Die dementsprechend durchgeführten LR- und Wald-Tests ergeben  $\chi^2(2)$ -Prüfgrößen von 5,47 und 5,46. Bei einer Irrtumswahrscheinlichkeit von 5% (kritischer Wert 5,99) kann die Nullhypothese eines Random Walks nicht abgelehnt werden. Dies bestätigt die mit Hilfe der ADF-Tests getroffene Diagnose bzgl. der Nichtstationarität der Zeitreihe. Da ein Random Walk nicht prognostizierbar ist, folgt, daß der Erklärungswert der Spezifikation bezüglich der Zinsänderung in der nächsten Periode nahe Null liegt.<sup>93</sup>

Da aus dem LEVEL-Modell, mit der Restriktion  $\gamma = 0,5$ , das in der Praxis weitverbreitete CIR-Modell folgt, wird geprüft, ob diese Parameterrestriktion abgelehnt werden kann. Die Prüfgrößen des LR- und des Wald-Tests sind  $\chi^2$  verteilt mit einem Freiheitsgrad. Beide Prüfgrößen bleiben mit berechneten Werten von 1,18 und 1,36 deutlich unter dem kritischen Wert von 3,84. Die Nullhypothese der Gültigkeit des CIR-Modells kann demnach nicht abgelehnt werden.

Die Ergebnisse der Schätzung mit dieser Restriktion sind in der Spalte CIR der Tabelle 2 dargestellt. Der Volatilitätsparameter  $\psi^2$  wird, sobald  $\gamma$  fixiert wird, hoch signifikant. Auch für das CIR Modell zeigen die D-Tests weder einen Strukturbruch noch eine Fehlspezifikation an. Auch wenn diese Tests keinen Strukturbruch anzeigen, muß

---

<sup>90</sup> Dieser Test testet die Hypothese  $\theta_{1/75-7/86} = \theta_{8/86-2/98}$  mittels des in Abschnitt 2.3.3 beschriebenen D-Tests. Da für den August 86 kein Strukturbruch vermutet wird, wäre eine Verwerfung dieser Nullhypothese ein Indiz für eine Fehlspezifikation.

<sup>91</sup> Für amerikanische Zinsdaten zwischen 1960 und 1993 kann dieses Modell ca. 20% der Volatilität in den Zinsänderungen prognostizieren. Vgl. Chan et al., 1990, S.1218 und Brenner et al., 1996, S.96.

<sup>92</sup> Ergebnisse sind hier nicht aufgeführt.

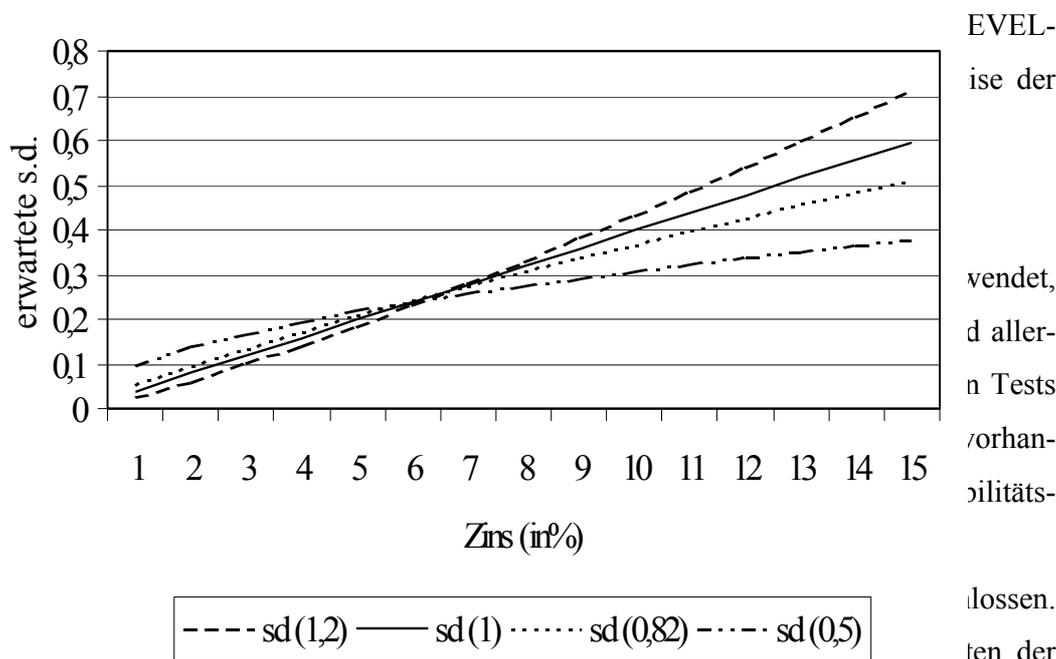
<sup>93</sup> Entsprechende  $R^2$  (analog zum  $R^2$  bzgl. der Volatilität) wurden berechnet. Diese liegen für alle Spezifikationen nur unwesentlich über Null.

vermutet werden, daß die höhere Volatilität vor 1975 nicht nur durch ein höheres Zinsniveau zu erklären ist.

Die CIR Spezifikation wird deshalb um eine Dummy Variable für den Volatilitätsparameter  $\psi^2$  für den Zeitraum nach 1975 erweitert (CIR-Dummy75). Diese wird negativ und hoch signifikant geschätzt. Der Volatilitätsparameter  $\psi^2$  wird für den Zeitraum bis 1975 ca. 5 x größer und für den Zeitraum danach ca. halb so groß geschätzt wie  $\psi^2$  im CIR-Modell ohne Dummy Variable. Durch diese Modifikation erhöht sich die Prognosegüte des CIR-Modells deutlich auf 10%.

Das CIR-Modell impliziert  $\chi^2$ -verteilte Zinsänderungen. Aus der Restriktion  $\gamma = 1$  hingegen folgt ein log-normalverteilter Zinsprozeß. Dies ist die Spezifikation nach Brennan und Schwartz.<sup>94</sup> Auch diese Restriktion kann nicht verworfen werden. Es muß festgestellt werden, daß diese Teststrategie nicht dazu geeignet ist, zwischen Modellen zu diskriminieren, die sich lediglich in ihrem Wert bzgl.  $\gamma$  unterscheiden. Die Ursache dieses Ergebnisses liegt in der Tatsache, daß die deutschen Zinsdaten überwiegend in Bereichen liegen, in denen verschiedene Werte von  $\gamma$  nicht zu signifikant unterschiedlichen Volatilitätserwartungen führen. Abbildung 1 verdeutlicht, daß sich solche Unterschiede in den Erwartungen erst ab Zinssätzen > 10% einstellen.

Aufgrund dieser Ergebnisse erscheint es sinnvoll Modelle zu untersuchen die zeitlich



GMM untersucht wurden, konzentriert sich die Forschung gegenwärtig auf die Frage, wie der Effizienznachteil der GMM gegenüber anderen Schätzmethoden, z.B. der Maximum Likelihood Methode, überwunden werden kann. Entscheidend scheint hierbei die Wahl der Momentbedingungen zu sein und so wurden in den letzten Jahren neue

<sup>94</sup> Vgl. Chan et al., 1992, S. 1212.

und systematische Methoden der Generierung von Momentbedingungen entwickelt, um die bislang eher willkürlichen Verfahren zu systematisieren.

Die Vorgehensweise bei einer GMM-Schätzung wurde anhand einer Schätzung des stochastischen Zinsprozesses verdeutlicht.

## Anhang A

### Bestimmung der Schätzer-Bandbreite

Bei dem von Newey und West benutzten Bartlett-Fenster muß  $m$  mit der Rate  $T^{1/3}$  steigen, um die größtmögliche Konvergenzgeschwindigkeit an die wahre Varianz-Kovarianzmatrix zu gewährleisten.<sup>95</sup> Folglich gilt für die optimale Bandbreite des Schätzers für  $S$ :  $m = \pi T^{1/3}$ . Es gilt den Parameter  $\pi$  so zu bestimmen, daß eine festzulegende Verlustfunktion bzgl. der Schätzerverzerrung und -varianz minimiert wird. Nimmt man für den Augenblick an, daß der Instrumentvektor  $z_t$  aus nur einem Element besteht, dann wird  $S$  zu einem Skalar und das Kriterium, mit dessen Hilfe das optimale  $\pi$  bestimmt wird, lautet  $(b^2 + v)$ . Mit  $b$  als der Verzerrung des Schätzers und  $v$  der Varianz, gilt es diesen Wert zu minimieren (MSE Kriterium).<sup>96</sup> Ist  $r$  allerdings größer als 1, dann wird das MSE Kriterium auf den Skalar  $w'(\bar{S} - S)w$  angewandt.  $\bar{S}$  ist der Schätzer und  $S$  die wahre Matrix. Der Vektor  $w$  fungiert hier als ein Gewichtungsfaktor, der das Kriterium auf eine skalare Größe transformiert. Für unterschiedliche Gewichtungsvektoren  $w$  gelangt man zu verschiedenen optimalen Parameterwerten. Sowohl Newey/West als auch Andrews/Monahan benutzen den  $(r \times 1)$  Vektor  $w = (0 \ 1 \ \dots \ 1)'$ , der den Elementen der Varianz-Kovarianzmatrix  $S$ , die in Verbindung mit dem ersten Instrument, der Konstanten, entstehen kein Gewicht gibt. Dieses  $w$  gilt für den Fall, daß das Modell eine Momentbedingung vorgibt. Im später behandelten Beispiel wurden zwei vorgegeben und der Gewichtungsvektor  $w$  wird zu einem  $(2r + 1)$  Vektor mit 0 an den Stellen 1 und  $r+1$  und 1 an den sonstigen Stellen. Unter diesen Bedingungen läßt sich das optimale  $\pi$  wie folgt schätzen:

$$\hat{\pi} = 1,1447(\{\hat{s}^{(1)}/\hat{s}^{(0)}\}^2)^{1/3} \text{ mit } \hat{s}^{(0)} = \sigma_0 + 2 \sum_{p=1}^n \hat{\sigma}_p \text{ und } \hat{s}^{(1)} = 2 \sum_{p=1}^n v \hat{\sigma}_p$$

sowie  $n = [4(T/100)^{2/9}]^{97}$ . Die Parameter  $\sigma_v$  werden mit Hilfe von

$$\hat{\sigma}_p = \frac{1}{T-1} \sum_{t=p+2}^T (w'v_t)(w'v_{t-p})$$

geschätzt, wobei  $v_t$  für den Vektor der geschätzten Residuen aus dem *pre-whitening* Verfahren steht. Kommt dieses Verfahren nicht zur Anwendung, so sind die  $v_t$ s durch die Residuen der Orthogonalitätsbedingungen  $\varepsilon_t$  zu ersetzen. Werden die Vektoren für  $w$  und  $v_t$  eingesetzt, erhält man

<sup>95</sup> Bei vielen anderen Fenstern (kernels) ist diese Rate  $T^{1/5}$ . Vgl. Newey/West, 1994, S.633.

<sup>96</sup> Vgl. Newey/West, 1994, S.633.

<sup>97</sup> Bei der Aufsummierung wird der ganzzahlige Teil von  $n$  benutzt. Der wahre Wert der Parameter summiert sich über  $v = 1 \dots \infty$  auf.  $n$  dient als Vereinfachung für die Schätzung ist aber lt. Newey und West (1994, S.635) kein kritischer Parameter für die Eigenschaften des Schätzers für  $S$ . Vgl. Newey/West, 1994, S.635.

$$\hat{\sigma}_v = \frac{1}{T-1} \sum_{t=p+2}^T (v_{2t} + v_{3t} + \dots + v_{rt})(v_{2t-p} + v_{3t-p} + \dots + v_{rt-p}).$$

Bei der Bestimmung der  $\hat{\sigma}_v$ s werden alle Elemente der Autokovarianzfunktion der Residuen  $v_t$  und  $v_{t-p}$  aufsummiert, ausgenommen Elemente derjenigen Reihen und Spalten, die mit der Konstanten als Instrument verbunden sind. Mit den berechneten  $\hat{\sigma}_v$ s für  $p = 1 \dots n$  kann man  $\hat{s}^{(0)}$ ,  $\hat{s}^{(1)}$ ,  $\hat{\pi}$  sowie  $m$  berechnen:  $m = \hat{\pi} T^{1/3}$ . Dieses  $m$  geht dann als Lag-Obergrenze in die Schätzung von  $S$  ein.

## Anhang B

### Pre-Whitening

Die Residuen, die dieser Transformation zu unterziehen sind, sind die  $\varepsilon_t$ s, die sich aus  $\varepsilon_t = f(\theta, x_t, z_t) = u_t z_t = u_t (1 \quad z_{1t} \quad z_{2t} \quad \dots \quad z_{rt})'$  ergeben. Mittels eines Vektor Autoregressiven (VAR) Ansatzes werden Residuen erzeugt, die weniger Autokorrelation aufweisen als die Residuen aus den Orthogonalitätsbedingungen  $\varepsilon_t$ . Diese neu entstandene Residuenreihe wird benutzt, um  $S$  zu schätzen.<sup>98</sup> Andrews und Monahan weisen nach, daß der resultierende Schätzer für  $S$  auch dann konsistent ist, wenn der spezifizierte AR-Prozeß nicht der wahre Prozeß ist. Die aus dieser Modellierung resultierenden Residuen sind dann diejenigen, mit deren Hilfe die Matrix  $S$  geschätzt wird. Newey und West folgend wird von einem AR(1)-Prozeß ausgegangen, der sich folgendermaßen darstellt:<sup>99</sup>

$$\varepsilon_t = \hat{A} \varepsilon_{t-1} + v_t.$$

$\hat{A}$  ist die  $(r \times r)$  Matrix der Koeffizientenschätzer,  $\varepsilon_t$  der wie oben definierte  $(r \times 1)$  Vektor der Residuen und  $v_t$  der  $(r \times 1)$  Vektor der aus der VAR-Spezifikation resultierenden Residuen.

Zur Verdeutlichung eines solchen VAR-Ansatzes folgt eine Darstellung in Matrizenform. Das erste Element des Instrumentvektors ist wieder die Konstante 1 und  $\hat{A}$  sei die  $(r \times r)$  Matrix der geschätzten Koeffizienten aus dem VAR-Ansatz.

<sup>98</sup> Vgl. Andrews/Monahan, 1992, S.954.

<sup>99</sup> Vgl. Newey/West, 1994, S. 634.

$$\varepsilon_t = \hat{A} \varepsilon_{t-1} + v_t$$

$$\Leftrightarrow \begin{pmatrix} u_t \\ u_t z_{1t} \\ u_t z_{2t} \end{pmatrix} = \hat{A} \begin{pmatrix} u_{t-1} \\ u_{t-1} z_{1t-1} \\ u_{t-1} z_{2t-1} \end{pmatrix} + v_t = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} u_{t-1} \\ u_{t-1} z_{1t-1} \\ u_{t-1} z_{2t-1} \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \\ v_{3t} \end{pmatrix}$$

Der Vektor der geschätzten Residuen  $v_t$  wird anstelle der  $\varepsilon_t$ s benutzt, um den Newey-West-Schätzer zu berechnen. pw steht für pre-whitened.

$$\hat{S}_T^{pw} = \hat{\Pi}_{0,T} + \sum_{v=1}^m (1 - (v/(m+1))) (\hat{\Pi}_{v,T} + \hat{\Pi}_{v,T}')$$

$$\text{mit } \hat{\Pi}_{v,T} = \frac{1}{T} \sum_{t=v+1}^T v_t v_{t-v}'.$$

Bevor die Inverse dieser Matrix als Gewichtungsmatrix in die GMM-Schätzung eingeht, muß sie einer zweiten Transformation unterzogen werden, die ihr die Informationen wieder zurück gibt, die bei der VAR-Schätzung von der Koeffizientenmatrix  $\hat{A}$  aufgenommen wurden.

$$\hat{S}_T^{NWpw} = (I_r - \hat{A})^{-1} \hat{S}_T^{pw} (I_r - \hat{A})^{-1'}$$

Die Inverse dieser Matrix ist dann die in der GMM Schätzung verwendete Gewichtungsmatrix.

Bei der praktischen Umsetzung dieser Datentransformation bleibt noch die Frage zu beantworten, wie die Koeffizientenmatrix  $A$  und die Residuen geschätzt werden. Das Standardverfahren zur Schätzung eines VAR Modells ist das Kleinstquadratverfahren.<sup>100</sup>

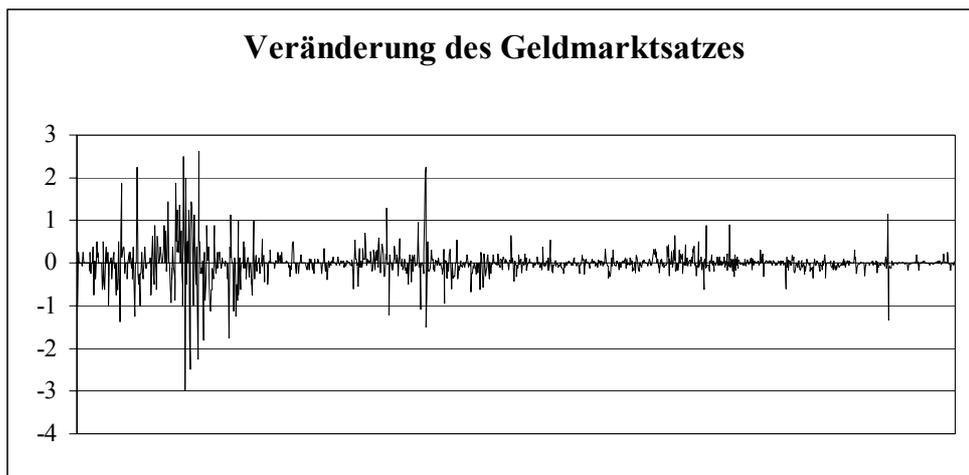
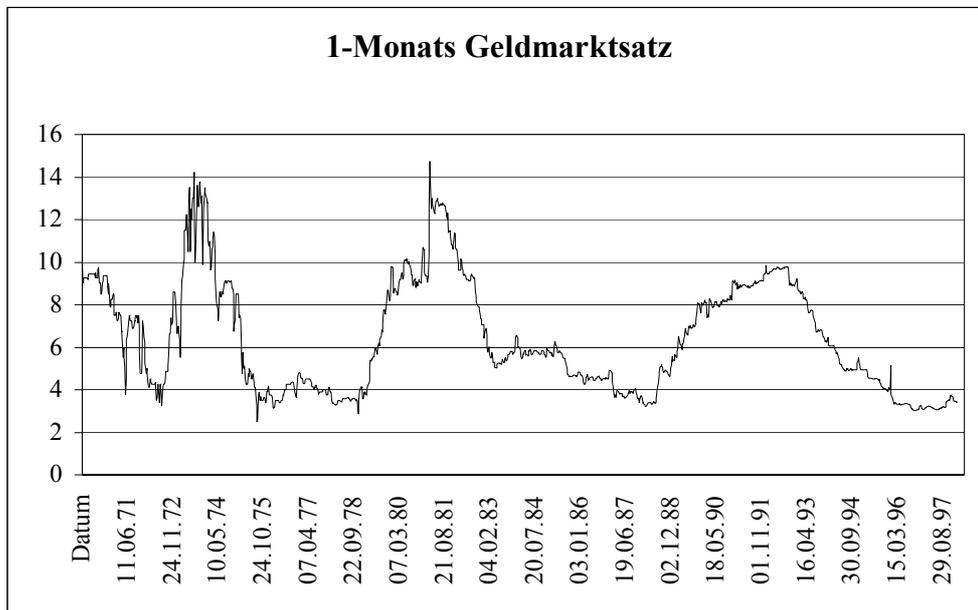
---

<sup>100</sup> Vgl. Maddala, 1992, S.578-581.

## Anhang C

### 1-Monats Geldmarktzins, 2.1.1970 bis 20.2.1998

wöchentliche Observationen



## Literaturverzeichnis

- Aït-Sahalia, Yacine**, 1996, Testing Continuous-Time Models of the Spot Interest Rate, in: *The Review of Financial Studies*, 1996, Vol. 9, S. 385-426.
- Andersen, Torben G., Lund, Jesper**, 1997, Estimating continuous-time stochastic volatility models of the short-term interest rate, in: *Journal of econometrics*, Vol. 77, S. 343-377.
- Andersen, Torben G., Sørensen, Bent E.**, 1996, GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, in: *Journal of Business & Economic Statistics*, Vol. 14, S.328-352.
- Andrews, Donald W.K., Fair, Ray C.**, 1988, Inference in Nonlinear Models with Structural Change, in: *Review of Economic Studies*, Vol. 55, S.615-639.
- Andrews, Donald W.K., Monahan, J.C.**, 1992, An Improved Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimator, in: *Econometrica*, Vol. 60, S. 953-966.
- Baxter, Martin, Rennie, Andrew**, 1996, *Financial Calculus - An introduction to derivative pricing*, Cambridge.
- Brenner, R.J., Harjes, R.H., Kroner, K.F.**, 1996, Another Look at Models of the Short-Term Interest Rate, in: *Journal of Financial and Quantitative Analysis*, Vol. 31, S. 85-107.
- Campbell, John Y., Lo, Andrew W., MacKinlay, A. Craig**, 1997, *The Econometrics of Financial Markets*, Princeton.
- Chan, K.C., Karolyi, G.A., Longstaff, F.A. und Sanders, A.B.**, 1992, An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, in: *The Journal of Finance*, Vol. 47, S.1209-1227.
- Chang, Alpha C.**, 1984, *Fundamental Methods of Mathematical Economics*, 3<sup>rd</sup> ed., Singapur.
- Davidson, Russel, MacKinnon, James G.**, 1993, *Estimation and Inference in Econometrics*, Oxford.
- Gallant, A. Ronald, Tauchen, George**, 1996, Which Moments to Match?, in: *Econometric Theory*, Vol. 12, S.657-681.
- Ghysels, Eric, Hall, Alastair**, 1990, Are Consumption-Based Intertemporal Capital Asset Pricing Models Structural?, in: *Journal of Econometrics*, Vol. 45, S.121-139.
- Härtter, Erich**, 1987, *Wahrscheinlichkeitsrechnung, Statistik und mathematische Grundlagen*, Göttingen.
- Hall, Alastair**, 1993, Some Aspects of Generalized Method of Moments Estimation, in: Maddala, G.S., Rao, C.R., Vinod, H.D. (eds.), *Handbook of Statistics*, Vol. 11, S.393-417.
- Hamilton, James D.**, 1994, *Time Series Analysis*, Princeton.
- Hansen, Lars Peter**, 1982, Large Sample Properties of Generalized Method of Moments Estimators, in *Econometrica*, Vol. 50, S.1029-1054.
- Hansen, Lars Peter, Scheinkman, José Alexandre**, 1995, Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes, in: *Econometrica*, Vol. 63, S.767-804.

- Hansen, Lars Peter, Singleton, Kenneth J.**, 1982, Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models, in *Econometrica*, Vol.50, S. 1269-1286.
- Hartung, Joachim**, 1986, Statistik, 5. Auflage, München.
- Haverkamp, Tom**, 1993, Ein Zweifaktormodell der Zinsstruktur: empirische Analyse und Bewertung zinsderivater Finanzinstrumente, Dissertation Hochschule St. Gallen, Bern.
- Hull, John C.**, 1991, Options, Futures and other Derivatives, 1<sup>st</sup> ed., London.
- Johnston, J., DiNardo, J.**, 1997, Econometric Methods, 4<sup>th</sup> ed., New York.
- Kmenta, Jan**, 1986, Elements of Econometrics, 2<sup>nd</sup> ed., New York.
- Nerlove, Marc, Grether, David M., Carvalho, José L.**, 1995, Analysis of Economic Time Series, San Diego.
- Newey, Whitney K.**, 1985, Generalized Method of Moments Specification Testing, in: *Journal of Econometrics*, Vol. 29, S. 229-256.
- Newey, Whitney K., West, D. Kenneth**, 1987a, A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, in: *Econometrica*, Vol. 55, S. 703-708.
- Newey, Whitney K., West, D. Kenneth**, 1987b, Hypothesis Testing with Efficient Method of Moments Estimation, in: *International Economic Review*, Vol. 28, S. 777-787.
- Newey, Whitney K., West, D. Kenneth**, 1994, Automatic Lag Selection in Covariance Matrix Estimation, in: *Review of Economic Studies*, Vol. 61, S. 631-653.
- Ogaki, Masao**, 1993a, Generalized Method of Moments: Econometric Applications, in: Maddala, G.S., Rao, C.R., Vinod, H.D. (eds.), *Handbook of Statistics*, Vol. 11, S. 455-488.
- Ogaki, Masao**, 1993b, GMM: A Users Guide, Working Paper No. 348, Rochester Center for Economic Research.
- Rebonato, Riccardo**, 1996, Interest-Rate Option Models, Chichester.
- Schervish, Mark J.**, 1995, Theory of Statistics, Springer Series in Statistics, New York.
- Tauchen, George**, 1986, Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data, in: *Journal of Business & Economic Statistics*, Vol. 4., S.397-416.

Autor:

Ralf Becker, Dipl.-Vwl., Projektmitarbeiter

Bisher erschienene Arbeitspapiere:

1. Peter M. Schulze, Prognoseverfahren wissenschaftlicher Institute in der Bundesrepublik Deutschland. Überblick über eine Umfrage (Dezember 1993)
2. Martina Nold / Peter M. Schulze, Möglichkeiten und Grenzen der Quantifizierung der Schattenwirtschaft (April 1994)
3. Armin Seher, Einfluß der Integrationsordnung bei Zeitreihen auf die Spezifikation von Fehlerkorrekturmodellen (Juni 1994)
4. Lars Berg / Armin Gemünden / Frank Hubert / Ralf Leonhardt / Michael Leroudier, Die Situation der Studentenschaft in den Wirtschaftswissenschaften an der Universität Mainz im Frühjahr 1994. Ergebnisse einer Umfrage (August 1994)
5. Christoph Balz, Ein Fehlerkorrekturmodell zur Entwicklung des Kapitalmarktzinses in der Bundesrepublik Deutschland (Oktober 1994)
6. Reinhard Elkmann / Nora Lauterbach / Stephan Wind, Tertiärisierung regionaler Wirtschaftsstrukturen. Eine empirische Analyse kreisfreier Städte und Landkreise in Hessen, Rheinland-Pfalz und dem Saarland (Dezember 1994)
7. Peter M. Schulze / Uwe Spieker, Deutsche Aktienindizes. Statistische Konzepte und Beispiele. (Dezember 1994)
8. Armin Seher / Peter M. Schulze, Fehlerkorrekturmodelle und die Bewertung von Aktienkursindizes. Empirische Analyse zur Eignung des Konzepts (Januar 1995)
9. Reinhard Elkmann / Annette Klostermann / Kerstin Lieder, Zur intertemporalen Konstanz der Struktur regionaler Lohn- und Gehaltsniveaus in der Bundesrepublik Deutschland (Mai 1995)

10. Christoph Fischer, Ein Fehlerkorrekturmodell zur Kaufkraftparitätentheorie (März 1996)
11. Ralf Becker / Claudia Müller, Zur Schätzung regionaler Konsumfunktionen (Oktober 1996)
12. Frank Hubert, Klassifizierung der Arbeitsmärkte in den OECD-Ländern mittels Cluster- und Diskriminanzanalyse (April 1997)
13. Frank Hubert, Das Okun'sche Gesetz: Eine empirische Überprüfung für ausgewählte OECD-Länder unter besonderer Berücksichtigung der nationalen Arbeitsmarktordnungen (September 1997)
14. Christoph Balz/ Peter M. Schulze, Die Rolle nationaler, regionaler und sektoraler Faktoren für die Variation von Output, Beschäftigung und Produktivität in der Bundesrepublik Deutschland (Dezember 1997)
15. Peter M. Schulze, Steigende Skalenerträge und regionales Wachstum: Eine quantitative Analyse mit kleinräumigen Daten (März 1998)
16. Ralf Becker, Die Verallgemeinerte Momentenmethode (Generalized Method of Moments - GMM). Darstellung und Anwendung (Juni 1998)