

TREND IN ÖKONOMETRISCHEN MODELLEN

Eine Untersuchung der statistischen
Konsequenzen für ausgewählte
Schätzverfahren

Walter KRÄMER

Forschungsbericht/
Research Memorandum No. 196

März 1984

Die in diesem Forschungsbericht getroffenen Aussagen liegen im Verantwortungsbereich des Autors und sollen daher nicht als Aussagen des Instituts für Höhere Studien wiedergegeben werden.

Inhaltsverzeichnis

1.	EINLEITUNG	3
2.	VORBEMERKUNGEN	9
	a) einige technische Lemmata	9
	b) Diskussion	23
3.	TREND BEI STANDARDREGRESSION UND KORRELIERTEN STÖRGRÖSSEN	27
	a) bivariate Regression mit linearem Trend	27
	b) bivariate Regression mit exogenen ARIMA-Vari- ablen	35
	c) Simulation	46
	d) Diskussion	55
4.	TREND BEI FEHLERN IN DEN VARIABLEN	59
	a) Modell und Schätzverfahren	59
	b) Konsistenz und Effizienz von OLS	69
	c) Simulation	79
	d) Diskussion	93
5.	TREND IN SIMULTANEN GLEICHUNGEN	95
	a) Modell und Schätzverfahren	95
	b) Konsistenz und Effizienz von OLS	100
	c) Simulation	110
	d) Diskussion	129
6.	ZUSAMMENFASSUNG	133
	ANHANG	137
	a) empirische Analyse der Korrelation ausgewählter ökonomischer Zeitreihen	139
	b) relative Effizienz des Erste-Differenzen Schätzers bei wachsender Störgrößenkorrelation	149
	LITERATURVERZEICHNIS	159

1. EINLEITUNG

Trend ist ein nahezu universelles Phänomen ökonomischer Zeitreihendaten. Versuche, ihn zu modellieren, zu schätzen und zu entfernen beschäftigen Ökonometriker und Statistiker seit den Anfängen ihrer Wissenschaft (in der Ökonometrie etwa FRISCH und WAUGH, 1933; TINTNER, 1940; 1960, Kap. 7; ANDERSON, 1971, Kap. 2; oder als Beispiele jüngerer Beiträge, CHAN et al., 1977, und NERLOVE et al., 1979, Kap. 12. Eine nützliche historische Übersicht findet man in NERLOVE et al., 1979, Kap. 1).

Nur wenige ökonomische Variable wie etwa Zinssätze oder Arbeitslosenraten weisen eine sachlogisch vorgegebene Obergrenze auf (und sogar hier verschwindet durch Logarithmieren jede natürliche Grenze, wenn auch nach unten). Man beobachtet im Gegenteil oft ein mehr oder weniger regelmäßiges Wachstum der ersten oder zweiten Stichprobenmomente im Zeitablauf. In der empirischen Ökonometrie wird das seit langem anerkannt, sei es durch Differenzenbildung zur Vermeidung von durch Trend induzierten Nonsenskorrelationen, sei es durch explizite Berücksichtigung einer Zeitvariablen, dem Prototyp eines linearen Trends, in makroökonomischen Modellen, oder durch Modellierung ökonomischer Zusammenhänge in Wachstumsraten statt in Originalvariablen. Von einer Mißachtung des Phänomens "Trend" in der Ökonometrie kann also keine Rede sein.

In seltsamem Gegensatz dazu steht die (implizite) Verleugnung von Trend, sobald nicht mehr die Zeitreihe selbst, sondern Parameterschätzungen eines ökonometrischen Modells und deren asymptotische Eigenschaften im Zentrum des Interesses stehen. Die dann nötigen

Annahmen zum Grenzverhalten der exogenen Modellvariablen schließen die Existenz von Trend fast regelmäßig aus.

Die wichtigste dieser Standardannahmen, die bis auf wenige später zu besprechende Ausnahmen der gesamten ökonometrischen Großstichproben-Asymptotik zugrundeliegt, fordert die Konvergenz oder zumindest Beschränktheit der Kreuzproduktmatrix der exogenen Modellvariablen, nach Normierung mit dem Stichprobenumfang. Diese Annahme ist so sehr Allgemeingut geworden, daß sie vielfach bei asymptotischen Aussagen nicht mehr explizit als Voraussetzung erscheint.

Die Existenz von Trend (im weiter unter präzisierten Sinn) impliziert jedoch, daß die zweiten empirischen Momente einer Variablen nicht beschränkt bleiben können.

Dieser offensichtliche Widerspruch zwischen der Akzeptierung von Trend in den Modellvariablen auf der einen und den Standardannahmen zum Grenzverhalten der Regressoren auf der anderen Seite läßt sich auflösen, indem man etwa asymptotische Aussagen als Grenzeigenschaften in einem (hypothetischen) Zufallsexperiment interpretiert, bei dem die aktuelle Stichprobe der exogenen Modellvariablen sukzessive wiederholt wird. In diesem Fall ist die Grenzmatrix der zweiten Stichprobenmomente identisch mit der aktuellen Momentenmatrix und damit trivialerweise beschränkt.

Alternativ dazu kann man aber auch nach den Konsequenzen fragen, wenn nicht Niveaugrößen (levels), sondern Differenzen oder Wachstumsraten das aktuell beobachtete Muster weiterführen. Dann ist die Koexistenz von Trend und herkömmlichen Standardannahmen ausgeschlossen, und es erscheint durchaus sinnvoll, den asymptotischen

Eigenschaften von Test- und Schätzverfahren auch in diesen Situationen nachzugehen.

Dazu will die vorliegende Arbeit einen Beitrag leisten. Ausgehend von der Existenz von Trend werden die daraus folgenden statistischen Konsequenzen für einige ausgewählte Modelle und Schätzverfahren untersucht.

Der Plan der Arbeit ist dabei wie folgt:

Im folgenden Kapitel 2 werden frühere Ansätze zur Berücksichtigung von Trend vorgestellt und einige in den weiteren Kapiteln benötigte Hilfsresultate abgeleitet. Kapitel 3 untersucht das bivariate Standardregressionsmodell, mit (potentiell) autokorrelierten Störgrößen. Hier wird geprüft, inwieweit Standardresultate zur asymptotischen Effizienz der Gewöhnlichen Kleinst-Quadrate Methode (im weiteren OLS für "Ordinary Least Squares") relativ zur Verallgemeinerten Kleinst-Quadrate Methode (im weiteren GLS für "Generalized Least Squares") und zum Erste-Differenzen Schätzer (im weiteren FD für "First Differences") auch bei Trend in der unabhängigen Variablen erhalten bleiben. Kapitel 4 untersucht die analoge Frage im Fehler-in-den-Variablen Modell, mit GLS und der Korregierten Kleinst-Quadrate Schätzung (im weiteren CLS für "Corrected Least Squares") als Vergleichsmaßstab, und Kapitel 5 verfolgt die Konsequenzen von Trend für die relative Effizienz von OLS und k-Klasse Schätzungen in Simultanen Gleichungen. Kapitel 6 faßt die Ergebnisse zusammen. In einem Anhang schließlich werden zwei für den Hauptteil zu umfangreiche Nebenbetrachtungen angestellt. Einmal wird an Hand konkreter ökonomischer Zeitreihendaten aus Österreich geprüft, inwieweit die in dieser Arbeit getroffenen Annahmen zum Zeitverlauf trendbe-

hafteter Variablen mit der Realität kompatibel sind, und zum anderen das Grenzverhalten von GLS und FD bei wachsender Störgrößenkorrelation diskutiert. Dies erscheint im vorliegenden Kontext deswegen interessant, weil wachsende Störgrößenkorrelation die bei Trend asymptotisch verschwindende relative Effizienz von FD unter Umständen konterkarieren kann.

Die Hauptkapitel 3 bis 5 folgen dabei einem einheitlichen Gliederungsschema: (i) Vorstellung von Modell und Schätzverfahren, (ii) analytische Ableitung asymptotischer Eigenschaften, (iii) Überprüfung der Relevanz der Aussagen aus (ii) für kleine Stichproben mit Hilfe einiger Monte Carlo Experimente, und (iv) Diskussion der relevanten Literatur und möglicher Verallgemeinerungen. Auch ohne Anspruch auf enzyklopädische Behandlung des Themas sind dabei die betrachteten Modelle hinreichend heterogen, daß eine bei den jeweiligen Grundlagen beginnende Darstellung der Problematik wenig sinnvoll erscheint und daher auch nicht angestrebt wurde.

Neben der Diskussion von Trend hält eine Betonung der relativen Qualitäten von OLS als zusätzlicher roter Faden diese Abhandlung zusammen. Es zeigt sich dabei insbesondere, daß bei Trend in den exogenen Variablen einige der aus der Standardanalys bekannten asymptotischen Defekte von OLS (etwa Inkonsistenz bei Fehlern in den Variablen oder Simultanen Gleichungen) unter Umständen verschwinden, und daß auch die asymptotische Verteilung mit derjenigen anderer Verfahren identisch ist. In diesem Sinn kann man den vorliegenden Beitrag daher in Fortführung einer früheren Arbeit (KRÄMER, 1980a) als Versuch einer Rehabilitation der Gewöhnlichen Kleinst-Quadrate Methode verstehen.

Verschiedene Teilergebnisse dieser Untersuchung wurden auf den Jahrestagungen 1981 und 1983 der Econometric Society in Amsterdam und Pisa, sowie der Jahrestagung 1982 der Deutschen Statistischen Gesellschaft in München und dem Econometric Society Winter Symposium 1983 in Blanmont/Belgien vorgetragen. Den Diskussions- teilnehmern dieser Veranstaltungen verdanke ich eine Reihe nütz- licher Hinweise und Verbesserungsvorschläge. Gleiches gilt für die anonymen Gutachter des Allgemeinen Statistischen Archivs sowie von Metrika, Econometrica, Economics Letters und Methods of Opera- tions Research, in denen mehr oder weniger vorläufige Resultate aus der vorliegenden Arbeit bereits erschienen sind oder demnächst erscheinen werden, und für die Teilnehmer am Ökonometrischen Forschungsseminar am Institut für Höhere Studien, insbesondere Manfred Deistler, Immanuel Bomze, Benno Pötscher und Robert Kunst. Wertvolle Ratschläge gaben auch Adrian Pagan, Roberto Mariano, Asatoshi Maeshiro, Roch Roy, Hans Schneeweiß, Arie van der Genugten und John Chipman. Klaus Plasser und Harald Sonnberger sowie die anderen Mitarbeiter der EDV-Gruppe am Institut für Höhere Studien schließlich haben mir stets hilfsbereit und mit großer Geduld bei den Monte Carlo Experimenten geholfen.

Hiermit danke ich nochmals ich nochmals allen, die zu dieser Arbeit beigetragen haben, für ihr Interesse und ihre Mitarbeit. Ver- bleibende Fehler und Unklarheiten sind allein dem Autor anzulasten.

2. VORBEMERKUNGEN

a) einige technische Lemmata

Ausgangspunkt der weiteren Überlegungen ist das Standard-
Regressionsmodell

$$y_t = x_t \beta + u_t, \quad t=1,2,\dots \quad (2.1)$$

oder in Matrixschreibweise (für einen Stichprobenumfang T)

$$y = X \beta + u. \quad (2.2)$$

Dabei steht $y = (y_1, \dots, y_T)'$ für den $T \times 1$ Beobachtungsvektor der abhängigen (endogenen, zu erklärenden) Variablen, und $\beta = (\beta_1, \dots, \beta_K)'$ für den unbekannt, zu schätzenden $K \times 1$ Vektor der Regressionskoeffizienten. $X = (x_{tk})_{t=1, \dots, T; k=1, \dots, K}$ ist die Beobachtungsmatrix der unabhängigen (exogenen, erklärenden) Variablen (Regressoren), und $u = (u_1, \dots, u_T)'$ ein $T \times 1$ Vektor unbekannter stochastischer Störgrößen.

Die Resultate dieses Kapitels basieren auf den folgenden Annahmen zu $\{x_t\}$ und $\{u_t\}$:

(i) $\{x_t\}$ und $\{u_t\}$ ($t=1,2,\dots$) sind auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, A, P) definiert (dies schließt die im weiteren hauptsächlich verwendeten nichtstochastischen Regressoren als Spezialfall ein).

(ii) es gibt ein $T_0 \in \mathbb{N}$ mit $\Pr(\text{Rang}(X)=K) = 1$ für $T=T_0$ (und damit für alle $T > T_0$). Die maximale derartige Indexmenge wird im weiteren T genannt.

(iii) x_t und u_s sind unabhängig für alle $t, s \in \mathbb{N}$. Insbesondere sind damit verzögerte endogene Variable unter den Regressoren ausgeschlossen.

(iv) die u_t ($t=1,2,\dots$) sind unabhängig und identisch verteilt, mit $E(u_t) = 0$ und $E(u_t^2) = \sigma^2$.

Für $T \in \mathcal{T}$ ist die OLS-Schätzung $\hat{\beta}$ für β damit gegeben durch

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (2.3)$$

Die Abhängigkeit von T bei $\hat{\beta}$ (und ebenso y , X und u) wird dabei im weiteren nicht explizit kenntlich gemacht, steht aber im Mittelpunkt des Interesses. Analog symbolisieren " \rightarrow ", " \xrightarrow{p} " und " \xrightarrow{D} " immer gewöhnliche, stochastische und Verteilungskonvergenz bei $T \rightarrow \infty$, falls nicht ausdrücklich anders gesagt.

Es sei zunächst das Grenzverhalten von $\hat{\beta}$ für den Fall nicht-stochastischer Regressoren untersucht. Insbesondere werden in den weiteren Kapiteln Resultate zur asymptotischen Verteilung (bei $T \rightarrow \infty$) von $\hat{\beta}$ benötigt. Die Standard-Textbuchanalyse dazu geschieht gewöhnlich unter der Annahme

$$\lim_{T \rightarrow \infty} T^{-1}X'X = Q \text{ (regulär)} \quad (2.4)$$

(siehe etwa THEIL (1971, S. 362, Satz 8.1), JOHNSTON (1972, S. 274))

oder DHRYMES (1970, S. 146)) und liefert unmittelbar

$$T^{1/2}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 Q^{-1}) . \quad (2.5)$$

Für die vorliegende Untersuchung ist (2.4) aber zu restriktiv. Insbesondere werden dadurch die im Zentrum der weiteren Überlegungen stehenden trendbehafteten exogenen Variablen ausgeschlossen. Man betrachte etwa den Fall $K=2$, $x_{t1} = 1$, $x_{t2} = t$ ($t=1,2,\dots$) als Prototyp einer Regressormatrix mit Trend. Hier gilt

$$X'X = \begin{bmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{bmatrix} , \quad (2.6)$$

d.h. $T^{-1}X'X$ konvergiert nicht. Allgemein gilt für einen polynomialen Trend p 'ter Ordnung $\sum_{t=1}^T x_{ti} = O(T^{2p+1})$, d.h. (2.4) ist notwendig verletzt.

KRÄMER(1984a) diskutiert die asymptotische Verteilung von OLS unter den allgemeineren Annahmen

$$\max_{1 \leq t \leq T} x_{tk}^2 / \sum_{t=1}^T x_{tk}^2 \longrightarrow 0 \quad (k=1, \dots, K) \quad (2.7)$$

und

$$D_T^{-1} X'X D_T^{-1} \longrightarrow Q^* \text{ (regulär)} \quad (2.8)$$

mit

$$D_T = \begin{bmatrix} (\sum_{t=1}^T x_{t1}^2)^{1/2} & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & (\sum_{t=1}^T x_{tK}^2)^{1/2} \end{bmatrix}$$

$$= \text{diag} (\|x_{\cdot 1}\| , \dots , \|x_{\cdot K}\|) , \quad (2.9)$$

$\| \cdot \|$ die euklidische Norm und $x_{\cdot i}$ als i 'tem Spaltenvektor von X .

GRENANDER (1954), ROSENBLATT (1956) und GRENANDER und ROSENBLATT (1957) diskutieren verwandte Fragen unter den Annahmen (2.8) ,

$$\sum_{t=1}^T x_{tk}^2 \longrightarrow \infty \quad (k = 1, \dots, K) \quad (2.10)$$

und

$$\sum_{t=1}^{T+n} x_{tk}^2 / \sum_{t=1}^T x_{tk}^2 \longrightarrow 1 \quad (n \in \mathbb{N}, k=1, \dots, K) \quad (2.11)$$

SCHNEEWEISS (1976) operiert unter Annahmen, die im vorliegenden Kontext wie folgt formulierbar sind:

$$T^{-1} H(T)^{-1} X' X H(T)^{-1} \longrightarrow \tilde{Q} \quad (\text{regulär}) \quad (2.12)$$

mit $H(T) = \text{diag}(h_1(T), \dots, h_K(T))$,

$$\lim_{T \rightarrow \infty} h_k(T) = 1 \text{ oder } \infty \quad (k=1, \dots, K) \quad (2.13)$$

und

$$h_k(T+1)/h_k(T) \longrightarrow 1 \quad (k=1, \dots, K) \quad (2.14)$$

EICKER (1963, 1967) und HUBER (1973) schließlich fordern nur, daß

$$\max_{1 \leq t \leq T} \gamma_{tt} \longrightarrow 0 \quad (2.15)$$

mit γ_{tt} aus $(\gamma_{st}) = X(X'X)^{-1}X'$.

Im weiteren seien (2.7) und (2.8) durch (K), (2.8), (2.10) und (2.11) durch (GR), (2.12), (2.13) und (2.14) durch (SCH) und (2.15) durch (E) symbolisiert, mit folgenden Zusammenhängen zwischen diesen Bedingungskomplexen:

SATZ (2.1): (K) und (GR) sind äquivalent, schwächer als (SCH) und stärker als (E), und alle Bedingungskomplexe sind schwächer (weniger restriktiv) als (2.4).

BEWEIS:

(i) (2.4) \Rightarrow (SCH) :

Dies folgt unmittelbar mit $h_k(T) = 1$ ($k=1, \dots, K; t=1, 2, \dots$). Da ferner X aus (2.6) (SCH) erfüllt (mit $h_1(T) = 1$, $h_2(T) = T$ und

$$\tilde{Q} = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}) ,$$

nicht aber (2.4), ist (2.4) echt restriktiver als (SCH) .

(ii) (SCH) \Rightarrow (GR) :

Wegen (2.12) gilt zunächst

$$T^{-1} h_i(T)^{-2} \sum_{t=1}^T x_{ti}^2 \longrightarrow c_i, \quad 0 < c_i < \infty, \quad i=1, \dots, K. \quad (2.16)$$

Wegen (2.13) gilt weiter $T^{-1} h_i(T)^{-2} \longrightarrow \infty$, damit wegen (2.16)

$\sum_{t=1}^T x_{ti}^2 \longrightarrow \infty$. Aus (2.16) folgt ferner, daß

$$(T+1)^{-1} h_i(T+1)^{-2} \sum_{t=1}^{T+1} x_{ti}^2 / T^{-1} h_i(T)^{-2} \sum_{t=1}^T x_{ti}^2 \longrightarrow 1 ,$$

damit wegen (2.14) aber auch

$$\sum_{t=1}^{T+1} x_{ti}^2 / \sum_{t=1}^T x_{ti}^2 \longrightarrow 1 ,$$

damit wiederum (2.11). Schließlich gilt

$$D_T^{-1} X'X D_T^{-1} = T^{1/2} D_T^{-1} H(T) \{T^{-1} H(T)^{-1} X'X H(T)^{-1}\} H(T) D_T^{-1} T^{1/2}$$

mit $T^{1/2} D_T^{-1} H(T) \rightarrow C = \text{diag}(c_1^{-1/2}, \dots, c_K^{-1/2})$ (regulär), d.h.

$$D_T^{-1} X'X D_T^{-1} \rightarrow C\tilde{Q}C \quad (\text{regulär}).$$

Daß (SCH) echt restriktiver ist als (GR), zeigt das Beispiel $K=1$, $x_t = t^{-1/2}$. Wegen $\sum_{t=1}^T x_t^2 = O(\ln(T))$ ist (GR) erfüllt, aber $T^{-1} X'X \rightarrow 0$. Für jede Folge $H(T)$ ($T=1,2,\dots$) mit

$$T^{-1} H(T)^{-1} X'X H(T)^{-1} \rightarrow \tilde{Q} \quad (\text{regulär, d.h. hier } 0 < \tilde{Q} < \infty)$$

gilt damit $H(T) \rightarrow 0$, d.h. (2.13) und damit (SCH) sind verletzt.

(iii) (GR) \Rightarrow (K):

Da (2.8) beiden Bedingungskomplexen gemeinsam ist, bleibt zu zeigen, daß (2.7) aus (2.10) und (2.11) folgt. Wegen (2.11) gilt aber

$$\sum_{t=1}^T x_{ti}^2 / \sum_{t=1}^{T-1} x_{ti}^2 = 1 + x_{Ti}^2 / \sum_{t=1}^{T-1} x_{ti}^2 \rightarrow 1$$

d.h. $x_{Ti}^2 / \sum_{t=1}^{T-1} x_{ti}^2 \rightarrow 0$, damit auch

$$x_{Ti}^2 / \sum_{t=1}^T x_{ti}^2 \rightarrow 0 \quad . \quad (2.17)$$

Sei $t^*(T)$ der Index, für den x_{ti}^2 ($1 \leq t \leq T$) das Maximum annimmt.

1. Fall: $\{t^*(T) \mid T=1,2,\dots\}$ ist endlich.

Dann gibt es ein $M \in \mathbb{R}$ mit $0 < M < \infty$ und $x_{t^*i}^2 < M$ für alle $t \in \mathbb{N}$, d.h.

$$\max_{1 \leq t \leq T} x_{ti}^2 / \sum_{t=1}^T x_{ti}^2 \leq M / \sum_{t=1}^T x_{ti}^2 \rightarrow 0 \quad (\text{wegen (2.10)}).$$

2. Fall: $\{t^*(T) \mid T=1,2,\dots\}$ ist unbeschränkt.

Dann gilt

$$\max_{1 \leq t \leq T} x_{ti}^2 / \sum_{t=1}^T x_{ti}^2 = x_{t^*(T),i}^2 / \sum_{t=1}^T x_{ti}^2 \leq x_{t^*,i}^2 / \sum_{t=1}^{t^*(T)} x_{ti}^2 ,$$

wobei die letzte Folge als unendliche Teilfolge von (2.17) ebenfalls gegen 0 streben muß.

(iv) (K) \Rightarrow (GR):

(2.10) folgt unmittelbar aus (2.7) . Eine weitere Konsequenz von (2.7) ist (2.17), damit auch

$$\sum_{t=1}^{T+1} x_{ti}^2 / x_{T+1,i}^2 = 1 + \sum_{t=1}^T x_{ti}^2 / x_{T+1,i}^2 \longrightarrow \infty \quad (2.18)$$

d.h. $x_{T+1,i}^2 / \sum_{t=1}^T x_{ti}^2 \longrightarrow 0$ und

$$\sum_{t=1}^{T+1} x_{ti}^2 / \sum_{t=1}^T x_{ti}^2 = 1 + x_{T+1,i}^2 / \sum_{t=1}^T x_{ti}^2 \longrightarrow 1 ,$$

woraus auch (2.11) sofort folgt.

(K) \Rightarrow (E):

Mit $x_t = (x_{t1}, \dots, x_{tK})$ gilt

$$Y_{tt} = x_t (X'X)^{-1} x_t' = x_t D_T^{-1} \{D_T (X'X)^{-1} D_T\} D_T^{-1} x_t' .$$

Wegen (2.8) bleibt die Matrix in geschweiften Klammern beschränkt, und aus (2.8) folgt $\max_{1 \leq t \leq T} x_t \cdot D_T^{-1} \longrightarrow 0$ und damit (2.15) .

Auf der anderen Seite ist (K) echt restriktiver als (E) , wie das folgende Gegenbeispiel zeigt: $K = 2$, $x_{t1} = 1$ ($t=1,2,\dots$),
 $x_{12} = -1$, $x_{22} = x_{32} = x_{42} = 1$ und

$$x_{t2} = (-1)^n , t > 4 , 4 \cdot 3^n < t < 4 \cdot 3^{(n+1)} , n \in \mathbb{N} .$$

Hier gilt

$$X'X = \begin{bmatrix} T & g(T) \\ g(T) & T \end{bmatrix} \quad (2.19)$$

mit $-T/2 \leq g(T) \leq T/2$ und

$$g(T) = -T/2 \quad \text{für } T = 4 \cdot 3^n \quad (n \text{ ungerade})$$

$$g(T) = T/2 \quad \text{für } T = 4 \cdot 3^n \quad (n \text{ gerade}) ,$$

$$\text{d.h. } D_T^{-1} X'X D_T^{-1} = \begin{bmatrix} 1 & g(T)/T \\ g(T)/T & 1 \end{bmatrix} \quad \text{konvergiert nicht.}$$

Auf der anderen Seite folgt aus (2.19) sofort $(X'X)^{-1} \rightarrow 0$,
damit wegen der Beschränktheit der x_{ti} ($t \in \mathbb{N}$, $i=1,2$) aber auch
(2.15). q.e.d.

Im weiteren wird Trend in der i 'ten exogenen Variablen
formell definiert durch

$$T^{-1} \sum_{t=1}^T x_{ti}^2 \rightarrow \infty , \quad (2.20)$$

eine durch die Standardannahme (2.4) ausgeschlossene, in den Annahmekomplexen (SCH), (K), (GR) und (E) aber zugelassene Situation. GRENANDER und ROSENBLATT (1957, S. 245-254) und EICKER (1967, S. 64-65) geben dazu weitere Beispiele an.

Auch unter der am wenigsten restriktiven Annahme (2.15) bleiben gewisse ökonomisch interessante Variablenverläufe weiterhin ausgeschlossen, etwa solche mit exponentiellem Trend. Wie das Beispiel $K=1$, $x_t=2^t$ zeigt, mit

$$\max_{1 \leq t \leq T} x_t^2 / \sum_{t=1}^T x_t^2 \longrightarrow 4 > 0 ,$$

ist hier (2.7) und damit wegen Satz (2.1) auch (2.15) verletzt.

Bei stochastischen Regressoren, wie weiter unten in Abschnitt 3b), wird (2.20) ersetzt durch

$$T^{-1} \sum_{t=1}^T x_{ti}^2 \xrightarrow{P} \infty , \quad (2.21)$$

und an die Stelle von (2.15) tritt

$$\max_{1 \leq t \leq T} \gamma_{tt} \xrightarrow{P} 0 . \quad (2.22)$$

In den weiteren Kapiteln werden an mehreren Stellen Grenzverteilungen von OLS-Schätzungen bei trendbehafteten exogenen Variablen benötigt. Dabei wird die asymptotische Normalität von OLS unter (2.15) (und damit auch unter (SCH), (K) und (GR)) durch EICKER (1967, S. 61, Satz 3.1) oder HUBER (1973, S. 803, Proposition 2.2) garantiert. Ein analoges Resultat auf der Basis von (2.22) existiert jedoch noch nicht und wird daher als nächstes gezeigt, ausgehend von folgendem Hilfsresultat:

LEMMA (2.2) (GÄNSSLER, STROBEL und STUTE, 1978, S. 207; siehe auch HALL und HEYDE, 1980, Abschnitt 3.2): Sei $\{S_{Tt} \mid 0 \leq t \leq T; T=T_0, T_0+1, \dots\}$ eine Dreiecksfolge quadratisch integrierbarer, reeller, auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, A, P) definierter Zufallsvariablen, mit $S_{Tt} = \sum_{i=0}^t Z_{Ti}$, $Z_{T0} = 0$, und sei $\{F_{Tt} \mid 0 \leq t \leq T, T=T_0, T_0+1, \dots\}$ eine Dreiecksfolge von Unter- σ -Algebren von A so daß S_{Tt} F_{Tt} -meßbar ist, F_{Tt} monoton aufsteigt in t für alle T , und $E(S_{Tt} \mid F_{T, t-1}) = S_{T, t-1}$. Dann gilt

$$S_{TT} \xrightarrow{D} N(0, \sigma^2) \quad (2.23)$$

falls

$$\sum_{t=1}^T E(Z_{Tt}^2 | F_{T,t-1}) \xrightarrow{P} \sigma^2 \quad \text{und} \quad (2.24)$$

$$\sum_{t=1}^T E(Z_{Tt}^2 I(Z_{Tt}^2 \geq \varepsilon) | F_{T,t-1}) \xrightarrow{P} 0 \quad (2.25)$$

für alle $\varepsilon > 0$, wobei $I(\cdot)$ die Indikatorfunktion des Ereignisses in Klammern bezeichnet.

Zum Beweis siehe die Originalliteratur. Lemma (2.2) liefert mit einigen Zusatzüberlegungen das gewünschte Resultat.

SATZ (2.3): Mit Bedingung (2.22) und unter den Voraussetzungen dieses Unterkapitels gilt

$$(X'X)^{1/2}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 I). \quad (2.26)$$

BEWEIS: Sei $a = (a_1, \dots, a_K)'$ ein $K \times 1$ Vektor nichtstochastischer Konstanten mit $\sum a_i = 1$. Dann reicht es zum Beweis von (2.26) zu zeigen, daß

$$a'(X'X)^{1/2}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2). \quad (2.27)$$

Für $T \geq T_0$ und $1 \leq t \leq T$ sei $W_{Tt} = a'(X'X)^{-1/2} x_t'$, $Z_{Tt} = W_{Tt} u_t$, $W_{T0} = Z_{T0} = 0$, und $S_{Tt} = \sum_{i=0}^t Z_{Ti}$. Dann ist S_{Tt} quadratisch integrierbar, mit $S_{TT} = a'(X'X)^{1/2}(\hat{\beta} - \beta)$. $F_{T0} = \langle W_{T0}, \dots, W_{TT} \rangle$ sei die von W_{T0}, \dots, W_{TT} erzeugte Unter- σ -Algebra von A , und analog seien $U_t = \langle u_1, \dots, u_t \rangle$ und $F_{Tt} = \langle F_{T0}, \dots, U_t \rangle$ definiert. Damit gilt $F_{T,t-1} \subseteq F_{Tt}$, u_t ist unabhängig von $F_{T,t-1}$ und S_{Tt} ist meßbar bezüglich F_{Tt} . Aus einigen wohlbekanntem Standardresultaten zu bedingten

Erwartungen (siehe etwa LAHA und ROHATGI, 1979, Satz 6.1.5 und 6.1.6)

folgt weiter

$$\begin{aligned} E(S_{Tt} | F_{T,t-1}) &= S_{T,t-1} + E(Z_{Tt} | F_{T,t-1}) \\ &= S_{T,t-1} + a'(X'X)^{-1/2} x_t' E(u_t | F_{T,t-1}) \\ &= S_{T,t-1} \end{aligned}$$

d.h. $\{S_{Tt}\}$ besitzt die in Lemma (2.2) geforderten Eigenschaften einer Martingal-Dreiecksfolge bezüglich $\{F_{Tt}\}$.

Analog weist man (2.24) nach:

$$\begin{aligned} \sum_{t=1}^T E(Z_{Tt}^2 | F_{T,t-1}) &= \sum_{t=1}^T a'(X'X)^{-1/2} x_t' x_t (X'X)^{-1/2} a E(u_t^2 | F_{T,t-1}) \\ &= \sigma^2, \end{aligned}$$

da $E(u_t^2 | F_{T,t-1}) = E(u_t^2) = \sigma^2$ und $\sum_{t=1}^T a'(X'X)^{-1/2} x_t' x_t (X'X)^{-1/2} a = 1$.

Zum Beweis von (2.25) betrachte man

$$\begin{aligned} \varepsilon > 0, \quad \sum_{t=1}^T E(Z_{Tt}^2 I(Z_{Tt}^2 > \varepsilon) | F_{T,t-1}) \\ &= \sum_{t=1}^T W_{Tt}^2 E(u_t^2 I(W_{Tt}^2 u_t^2 > \varepsilon) | F_{T,t-1}) \\ &\leq \sum_{t=1}^T W_{Tt}^2 E(u_t^2 I(u_t^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) | F_{T,t-1}) \\ &= \sum_{t=1}^T W_{Tt}^2 E(u_t^2 I(u_t^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) | F_{T_0}). \end{aligned}$$

Die letzte Zeile, in der $F_{T,t-1}$ durch F_{T_0} ersetzt wird, folgt aus

$F_{T,t-1} = \langle F_{T_0}, u_{t-1} \rangle$ und der Tatsache, daß sowohl F_{T_0} als auch

$u_t^2 I(u_t^2 > \varepsilon / \max W_{Tt}^2)$ von u_{t-1} unabhängig sind (LAHA und ROHATGI,

1979, Satz 6.1.7) . Wegen $F_{T0} = \langle W_{T0}, \dots, W_{TT} \rangle$ ist aber $E(u_t^2 I(u_t^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) | F_{T0})$ eine Funktion der W_{Tt} 's, die außerdem wegen der iid-Eigenschaft der u_t 's für alle $t=1, \dots, T$ den gleichen Wert annimmt. Daher kann sie vor das Summenzeichen gezogen werden. Wegen $\sum_{t=1}^T W_{Tt}^2 = 1$ folgt daraus aber für beliebiges t' ($1 \leq t' \leq T$)

$$\begin{aligned} & \sum_{t=1}^T E(Z_{Tt}^2 I(Z_{Tt}^2 > \varepsilon) | F_{T, t-1}) \\ & \leq E(u_{t'}^2, I(u_{t'}^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) | F_{T0}). \end{aligned}$$

Aber

$$\begin{aligned} & P\{ E(u_{t'}^2, I(u_{t'}^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) | F_{T0}) = 0 \} \\ & \geq P\{ u_{t'}^2, I(u_{t'}^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) = 0 \} \\ & \geq P\{ I(u_{t'}^2 > \varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2) = 0 \} \\ & \xrightarrow{P} 1. \end{aligned}$$

Dabei folgt die letzte Zeile aus $W_{Tt}^2 \leq \gamma_{tt}$ und damit (wegen (2.15)) $\varepsilon / \max_{1 \leq t \leq T} W_{Tt}^2 \xrightarrow{P} \infty$. q.e.d.

(2.26) weicht von der üblichen Formulierung von Grenzverteilungsaussagen insofern ab, als die Matrix $(X'X)^{1/2}$ (die für $T \in \mathcal{T}$ immer existiert) und nicht eine Folge $g(T)$ von skalaren Konstanten zur Standardisierung von $(\hat{\beta} - \beta)$ herangezogen wird. Wenn eine derartige Folge $\{g(T)\}_{T \in \mathcal{N}}$ existiert mit

$$g(T)^{-2} X'X \xrightarrow{P} Q \text{ (regulär) ,}$$

kann die Grenzverteilung von $g(T)(\hat{\beta} - \beta)$ unmittelbar aus derjenigen von

$(X'X)^{1/2}(\hat{\beta} - \beta)$ abgeleitet werden, und umgekehrt. Üblicherweise ist $g(T) = T^{1/2}$. Die Formulierung (2.26) erlaubt darüber hinaus auch Aussagen in Fällen, wo eine skalare Folge zur Standardisierung von $X'X$ (und damit auch $(\hat{\beta} - \beta)$) nicht ausreicht (wie im Fall (2.6)), $g(T)^{-2}X'X$ zwar beschränkt und regulär bleibt, aber nicht konvergiert (wie in (2.19)), oder $g(T)^{-2}X'X$ gegen eine nichtdegenerierte Zufallsmatrix statt einer Konstanten konvergiert (wie weiter unten in 3.b).

Die Kapitel 4 und 5 der vorliegenden Arbeit setzten die Bedingungen (K) (oder äquivalent (GR)) für die exogenen Variablen der jeweiligen Modelle voraus. In diesem Fall kann auch D_T aus Definition (2.9) zur Standardisierung von $\hat{\beta} - \beta$ verwendet werden. Satz (2.3) liefert hier unmittelbar

$$D_T(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma^2 Q^*{}^{-1}) . \quad (2.28)$$

Das folgt sofort aus $D_T(\hat{\beta} - \beta) = D_T(X'X)^{-1/2} \{ (X'X)^{1/2}(\hat{\beta} - \beta) \}$ und

$$D_T(X'X)^{-1/2} \longrightarrow M$$

mit $M'M = Q^*{}^{-1}$.

Um Satz (2.3) auf die übliche Weise zur statistischen Inferenz heranzuziehen, bleibt noch die Konsistenz der OLS-gestützten Schätzung für σ^2 zu zeigen.

SATZ (2.4): Mit Bedingung (2.22) und unter den Voraussetzungen dieses Unterkapitels gilt

$$s^2 = (\hat{y} - X\hat{\beta})'(\hat{y} - X\hat{\beta}) / (T-K) \xrightarrow{P} \sigma^2 .$$

BEWEIS: Wegen

$$s^2 = \frac{1}{T-K} u'u - \frac{1}{T-K} u'X(X'X)^{-1}X'u \quad (2.29)$$

und $u'u/(T-k) \xrightarrow{P} \sigma^2$ reicht es zu zeigen, daß der zweite Ausdruck in (2.29) gegen 0 geht für $T \rightarrow \infty$. Dies wiederum folgt wegen $u'X(X'X)^{-1}X'u \geq 0$ aus

$$E\left(\frac{1}{T-K} u'X(X'X)^{-1}X'u\right) \xrightarrow{P} 0 . \quad (2.30)$$

Mit $(\gamma_{st}) = X(X'X)^{-1}X'$ wie in (2.15) und wegen $\sum_{t=1}^T \gamma_{tt} = \text{Rang}(X(X'X)^{-1}X') = K$ gilt aber

$$\begin{aligned} E\left(\frac{1}{T-K} u'X(X'X)^{-1}X'u\right) &= E\left(\frac{1}{T-K} \sum_{s,t=1}^T u_s u_t \gamma_{st}\right) \\ &= \frac{1}{T-K} \sum_{s,t=1}^T E(u_s u_t) E(\gamma_{st}) \\ &= \frac{1}{T-K} \sigma^2 E\left(\sum_{t=1}^T \gamma_{tt}\right) \\ &= \sigma^2 \frac{K}{T-K} \end{aligned}$$

$\longrightarrow 0 ,$

q.e.d.

Satz (2.4) liefert auch die Konsistenz der alternativen Schätzung $\tilde{s}^2 = s^2(T-K)/T$. Die in der Lehrbuchliteratur übliche Annahme normalverteilter Störgrößen ist nicht notwendig, und wie der Beweis ferner zeigt, kann auch auf (2.22) verzichtet werden. Wegen

$$\sum_{s=1}^T \gamma_{st}^2 \leq \gamma_{tt}^2 \leq (\text{Spur}(X(X'X)^{-1}X'))^2 \leq K^2$$

existiert $E(\gamma_{st})$ unabhängig von der speziellen Gestalt von X , und mit einem verallgemeinerten Inversen an Stelle von $(X'X)^{-1}$ für den Fall $\text{Rang}(X) < K$ zeigt der Satz damit die Konsistenz von s^2 ohne jede Voraussetzung zu X und ist damit auch allgemeiner als DRYGAS (1973) oder KLOEK (1972).

b) Diskussion

Aus (2.22) folgt unter anderem $(X'X)^{-1} \xrightarrow{p} 0$, d.h. $\|x_{\cdot k}\| \xrightarrow{p} \infty$. Satz (2.3) liefert aber $\|x_{\cdot k}\| (\hat{\beta}_k - \beta_k) = o_p(1)$ ($k=1, \dots, K$), damit $\hat{\beta}_k - \beta_k = o_p(1)$ und die Konsistenz von OLS. Jedoch ist (2.22) dafür unnötig restriktiv. Bei nichtstochastischen Regressoren ist wegen $\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$

$$(X'X)^{-1} \longrightarrow 0 \tag{2.31}$$

notwendig und hinreichend für die Quadratmittel-Konsistenz von $\hat{\beta}$ (auch für unkorrelierte stattunabhängige Störgrößen), damit auch hinreichend

für schwache Konsistenz. (2.31) ist erheblich schwächer als (2.15) und läßt insbesondere exponentielles Wachstum zu.

DRYGAS (1976) zeigt auch die Notwendigkeit von (2.31) für schwache Konsistenz. Unter der Zusatzvoraussetzung normalverteilter Störgrößen zeigen ANDERSON und TAYLOR (1976), daß (2.31) auch starke Konsistenz impliziert. Ohne diese Zusatzvoraussetzung beweist DRYGAS (1976) starke Konsistenz für

$$(h(T))^{-1} X'X \longrightarrow Q \text{ (regulär)} \quad (2.32)$$

mit einer skalaren, nichtstochastischen Folge $\{h(T)\}$ und $h(T) \rightarrow \infty$ für $T \rightarrow \infty$. Für $K = 1$ ist (2.32) äquivalent zu (2.31), andernfalls restriktiver. LAI, ROBBINS und WEI (1978) schließlich weisen starke Konsistenz unter den gegenwärtigen Rahmenbedingungen schon unter (2.31) nach.

WILLERS (1978), ANDERSON und TAYLOR (1979), CHRISTOPEIT und HELMES (1979), SCHMID (1981) und EICKER (1982) schließlich diskutieren Konsistenzbedingungen auch bei stochastischen Regressoren, bei teilweiser Berücksichtigung von korrelierten oder heteroskedastischen Störgrößen. Je nach Störgrößenverhalten sind die resultierenden Bedingungen wiederum schwächer als (2.22), das stochastische Analogon zu (2.15).

Statt der Frage "Wie muß die Folge der Regressorenmatrizen beschaffen sein, damit $\text{plim } \hat{\beta} = \beta$ (bzw. $\hat{\beta} \rightarrow \beta$ f.s.) d.h. $\text{plim } c'\hat{\beta} = \beta$ für alle festen K -Vektoren c ?" kann man auch das Problem formulieren "Wie muß bei gegebenem K -Vektor c die Folge der Regres-

sorenmatrizen beschaffen sein, damit $\text{plim } c'\hat{\beta} = c'\beta$?" Das ist die Frage nach den konsistenten Richtungen von OLS und wird etwa von WU (1980) diskutiert.

Satz (2.3) beweist die asymptotische Normalität von OLS natürlich auch für konstante (als Spezialfall von stochastischen) Regressoren. In diesem Fall ist (2.15) auch notwendig für die asymptotische Normalität von OLS (HUBER, 1973, Proposition 2.2). Dagegen sind die hier getroffenen Voraussetzungen zu den Störgrößen möglicherweise unnötig restriktiv. EICKER (1967, Satz 3.1) beweist das Analogon von Satz (2.3) auch für nicht notwendig identisch verteilte und potentiell heteroskedastische Störgrößen. WHITE (1980) und VAN DER GENUGTEN (1982) kommen auch bei stochastischen Regressoren ohne identisch verteilte Störgrößen aus, benötigen aber Annahmen zu höheren Momenten. Außerdem wird Trend in exogenen Variablen explizit ausgeschlossen (WHITE), oder die weiter unten interessierenden Spezialfälle werden durch sonstige Restriktionen nicht erfaßt (VAN DER GENUGTEN). LAI und WEI (1982) schließlich lassen zwar Trend zu, sind aber wegen fehlender Unabhängigkeit von Regressoren und Störgrößen ebenfalls zum Ausschluß der hier vor allem interessierenden Regressorentypen gezwungen.

Wie aus EICKER (1967, Lemma 3.2) hervorgeht, ist (2.15) nur unerheblich schwächer als (K) und (GR). Insbesondere ist (2.7) notwendig für (2.15) und liefert damit eine leicht zu kontrollierende und zu interpretierende Bedingung: Das Quadrat jeder Beobachtung muß relativ zur Quadratsumme asymptotisch vernachlässigbar sein.

Die weiter unten in 3 b) auftretenden stochastischen Regressoren zeigen Trend durch (für $T \rightarrow \infty$) wachsende, aber bei festem T endliche Varianz, während der Erwartungswert konstant bleibt. Darüber hinaus ist (2.21) auch möglich für nicht quadratisch integrierbare, identisch verteilte x_{ti} ($t=1,2,\dots; i=1,\dots,K$). Derartige Situationen werden von KANTER und STEIGER (1974) oder YOHAI und MARONNA (1977) untersucht, bleiben aber im folgenden außer Betracht.

3. TREND BEI STANDARD-REGRESSION UND KORRELIERTEN STÖRGRÖSSEN

a) bivariate Regression mit linearem Trend

Im Mittelpunkt dieses einführenden und illustrativen Abschnitts steht das Modell

$$y_t = \beta_1 + \beta_2 t + u_t \quad (3.1)$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad (t=1, \dots, T) \quad (3.2)$$

mit $|\rho| < 1$, $E(u_t) = E(u_t \varepsilon_t) = E(u_t \varepsilon_s) = 0$ ($t < s$) und $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ (d.h. $\sigma^2 = E(u^2) = \sigma_\varepsilon^2 / (1 - \rho^2)$). Da hier nur das Grenzverhalten der Varianzen konkurrierender Schätzverfahren und nicht deren asymptotische Verteilungen interessiert, wird allein Unkorreliertheit (statt Unabhängigkeit) der ε_t 's vorausgesetzt, d.h. $E(\varepsilon_t) = E(\varepsilon_t \varepsilon_s) = 0$ ($t \neq s$; $1 \leq s, t \leq T$).

Trotz (oder wegen) seiner einfachen Gestalt war das durch (3.1) und (3.2) definierte Modell bisher Gegenstand zahlreicher ökonometrisch-methodischer Untersuchungen (PRAIS und WINSTEN, 1954; ROSENBLATT, 1956; HANNAN, 1960, Kapitel V; MAESHIRO, 1976; PARK und MITCHELL, 1980; CHIPMAN, 1979; KRÄMER, 1982). Die zugrundeliegende X-Matrix ist die gleiche wie in (2.6), und die einzige Abweichung vom Standard-Regressionsmodell sind die AR(1)-Störgrößen u_t .

Im Einvernehmen mit der Literatur werden im weiteren die relative Effizienz (gemessen durch den Quotienten der Varianzen) alternativer Schätzverfahren für den Trendkoeffizienten β_2 untersucht, und einige Mißverständnisse dazu ausgeräumt. Die Schätzverfahren sind OLS, die Verallgemeinerte Kleinst-Quadrate Methode (im weiteren GLS für "Generalized Least Squares"), der Erste-Differenzen Schätzer (FD für "First Differences") und die durch Anwendung der Cochrane-Orcutt Transformation definierte

Schätzfunktion COT . Die jeweiligen Schätzungen sind gegeben durch

(2.3) (OLS) sowie

$$\hat{\beta}^{\text{GLS}} = (X'V^{-1}X)^{-1}X'V^{-1}y \quad , \quad (3.3)$$

$$\hat{\beta}^{\text{FD}} = ((DX)'DX)^{-1}(DX)'Dy \quad , \quad (3.4)$$

$$\hat{\beta}^{\text{COT}} = (X*'X*)^{-1}X*'y^* \quad . \quad (3.5)$$

Dabei ist $V = (v_{ij}) = 1/\sigma^2 E(uu')$ die Korrelationsmatrix von u ,

mit

$$v_{ij} = \rho^{|i-j|} \quad (3.6)$$

und D die $(T-1) \times T$ -Differenzenmatrix (siehe auch A.4), sowie

$$y_t^* = y_t - \rho y_{t-1} \quad ; \quad x_t^* = x_t - \rho x_{t-1} \quad (t=2, \dots, T) \quad . \quad (3.7)$$

COT und FD sind äquivalent zu OLS , angewandt auf ein transformiertes und um eine Beobachtung vermindertes Modell. GLS unterscheidet sich bei stationären AR(1) Störgrößen von COT allein durch Hinzufügen von

$$y_1^* = (1-\rho^2)^{1/2} y_1 \quad , \quad x_1^* = (1-\rho^2)^{1/2} x_1 \quad . \quad (3.8)$$

Falls wie hier das Modell eine Konstante enthält, entfällt die entsprechende Spalte von DX (und der entsprechende Koeffizient von β) im transformierten Modell, d.h. mit FD sind nur Steigungskoeffizienten zu schätzen. Im vorliegenden Spezialfall gilt $DX = (1, 1, \dots, 1)'$, d.h.

$$\hat{\beta}_2^{(\text{FD})} = (y_T - y_1) / (T-1) \quad . \quad (3.9)$$

Der Autokorrelationskoeffizient ρ der Störgrößen wird im folgenden für GLS und FD als bekannt vorausgesetzt. Diese Methoden sind daher vor allem als Vergleichsmaßstab und kaum für praktische Anwendungen von Interesse. Als praktikable Approximation bei hoher positiver Störgrößenkorrelation dient oft FD (die auf geschätztem ρ beruhenden praktisch anwendbaren Varianten von GLS und COT bleiben im folgenden außer Betracht).

Alle hier zugelassenen Schätzverfahren für β_2 sind linear und erwartungstreu. Daher befaßt sich die weitere Diskussion allein mit den Varianzen als Gütekriterium.

SATZ (3.1): In dem durch (3.1) und (3.2) definierten Modell gilt

$$\begin{aligned} \text{var}(\hat{\beta}_2^{(OLS)}) &= 12\sigma_{\varepsilon}^2 \{ -6\rho^{T+1}((T-1)\rho - (T+1))^2 - (T^3 - T)\rho^4 \\ &\quad + 2(T^2 - 1)(T-3)\rho^3 + 12(T^2 + 1)\rho^2 \\ &\quad - 2(T^2 - 1)(T+3)\rho + (T^3 - T) \} / \\ &\quad (1 - \rho^2)(1 - \rho)^4(T^3 - T)^2, \end{aligned} \quad (3.10)$$

$$\text{var}(\hat{\beta}_2^{(COT)}) = 12\sigma_{\varepsilon}^2 / (1 - \rho)^2(T^3 - 3T^2 + 2T), \quad (3.11)$$

$$\text{var}(\hat{\beta}_2^{(FD)}) = 2\sigma_{\varepsilon}^2(1 - \rho^{T-1}) / (1 - \rho^2)(T-1)^2, \quad (3.12)$$

$$\begin{aligned} \text{var}(\hat{\beta}_2^{(GLS)}) &= 12\sigma_{\varepsilon}^2 / (T-1)((T-3)(T-2)\rho^2 \\ &\quad - 2(T-3)(T+1)\rho + T(T+1)) . \end{aligned} \quad (3.13)$$

BEWEIS: (3.10) und (3.13) folgen unmittelbar aus CHIPMAN (1979, S. 120, Formel (2.11) - (2.13)), und (3.12) folgt aus (3.9) und

$$\text{cov}(y_t, y_s) = \text{cov}(u_t, u_s) = \sigma_{\varepsilon}^2 \rho^{|t-s|} / (1-\rho^2) . \quad (3.14)$$

(3.11) ergibt sich analog durch triviale Umformung aus der Formel

$$\text{var}(\hat{\beta}_2^{(\text{COT})}) = \sigma^2 / \sum_{t=2}^T (t^* - \bar{t}^*)^2,$$

$$\text{mit } \bar{t}^* = T^{-1} \sum_{t=2}^T t^* .$$

q.e.d.

Tabelle (3.1) gibt für $T=20$ und verschiedene ρ die Werte dieser Varianzen wieder. In Fortführung einer langanhaltenden Diskussion in der Literatur, ob oder ob nicht die C-0 Transformation einen Effizienzverlust gegenüber FD und OLS bewirkt, sind alle Varianzen relativ zu COT angegeben. Dies erlaubt auch einen direkten Vergleich mit MAESHIRO (1976, S. 498, Tabelle 1) (die erste Zeile in Tabelle (3.1) ist identisch mit der Zeile "Bt" in Maeshiro).

TABELLE (3.1): Relative Effizienz der Schätzverfahren für $T=20$ und verschiedene ρ .

	ρ (rho)								
	-.98	-.80	.00	.20	.40	.60	.80	.90	.98
var(OLS)/var(COT)	11.50	1.44	.86	.80	.74	.63	.39	.17	.012
var(FD)/var(COT)	525.60	28.83	3.18	2.11	1.35	.79	.35	.14	.010
var(GLS)/var(COT)	.998	.98	.86	.80	.71	.57	.32	.14	.010

Die Tabelle dokumentiert einen beträchtlichen Effizienzvorteil von COT verglichen mit OLS und insbesondere verglichen mit FD für negative und (im Fall FD) mäßig positive ρ , der aber für $\rho \rightarrow 1$ verschwindet (GLS als BLUE (Best Linear Unbiased Estimator) ist natürlich

immer besser als COT) . Für festes T und σ_ε^2 folgt aus Satz (3.1) sofort, daß

$$\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{FD}}) = \lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{GLS}}) = \sigma_\varepsilon^2 / (T-1) \quad , \quad (3.15)$$

während $\text{var}(\hat{\beta}_2^{\text{COT}}) \rightarrow \infty$ (als Folge von $\lim_{\rho \rightarrow 1} \sigma^2 = \infty$; ebenso könnte man σ^2 bei variierendem ρ festhalten. Das ist hier unerheblich, da sich dieser Faktor beim Quotientenbilden herauskürzt).

Damit wird COT für wachsendes ρ , wie auch in Tabelle (3.1) zu sehen, zunehmend ineffizient. Wegen (3.15) ist FD dagegen approximativ BLUE für hohes ρ und keinesfalls äquivalent zum immer schlechter werdenden COT, wie oft fälschlich behauptet wird (KADIYALA, 1968, S. 93; MAESHIRO, 1976, S. 499). Wegen (3.7) gilt zwar $\lim_{\rho \rightarrow 1} t^* = (1, 1, \dots, 1)'$ (der Regressor bei FD), aber im Gegensatz zu FD wird COT durch Anwendung von OLS auf ein Modell mit einer Konstanten geschätzt, wobei die für $\rho \rightarrow 1$ zunehmende Kollinearität der Regressoren den Qualitätsabfall bei COT verursacht.

Aus (3.15) folgt weiterhin, daß $\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{GLS}}) / \text{var}(\hat{\beta}_2^{\text{COT}}) = 0$, d.h. eine positive untere Schranke für diesen Quotienten existiert nicht. Damit ist die von CHIPMAN (1979, S. 117) angegebene untere Schranke von .535898 offensichtlich falsch. Diese Schranke entstand offenbar durch Analogieschluß aus dem Modell $y_t = \beta t + u_t$ (d.h. einer homogenen Trend-Regression), unter Vernachlässigung der Tatsache, daß durch Hinzufügen einer Konstanten das Grenzverhalten für $\rho \rightarrow 1$ von COT (nicht dagegen von OLS, GLS und FD) drastisch verändert wird. Wie ferner von Chipman persönlich mitgeteilt, gibt .535898 auch in diesem einfachen Modell nur den $\lim_{T \rightarrow \infty} \min_{0 \leq \rho < 1} \text{var}(\hat{\beta}^{\text{GLS}}) / \text{var}(\hat{\beta}^{\text{COT}})$ an, während die untere Schranke für diesen Quotienten .5 beträgt und für $\rho = 0$ und $T=2, 3$

auch angenommen wird.

Man rechnet weiterhin leicht nach, daß für beliebiges (aber festes) T auch $\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{OLS}}) / \text{var}(\hat{\beta}_2^{\text{COT}}) = 0$. Das wird schon durch die Tabelle nahegelegt und weist COT bei gegebenem T als (ab einem $\rho(T)$) schlechtestes aller zur Debatte stehenden Verfahren aus.

Ein vollkommen verschiedenes Grenzverhalten der relativen Effizienzen erhält man aber für fixes ρ und $T \rightarrow \infty$. Wie unmittelbar aus Satz (3.1) zu sehen, gilt für beliebiges ρ ($-1 < \rho < 1$): $\text{var}(\hat{\beta}_2^{\text{OLS}}) = O(T^{-3})$, $\text{var}(\hat{\beta}_2^{\text{COT}}) = O(T^{-3})$, $\text{var}(\hat{\beta}_2^{\text{GLS}}) = O(T^{-3})$, während $\text{var}(\hat{\beta}_2^{\text{FD}}) = O(T^{-2})$. Daraus und aus (3.15) folgt sofort das bemerkenswerte Resultat, daß $\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{FD}}) / \text{var}(\hat{\beta}_2^{\text{COT}}) = 0$ für jedes feste T , und $\lim_{T \rightarrow \infty} \text{var}(\hat{\beta}_2^{\text{FD}}) / \text{var}(\hat{\beta}_2^{\text{COT}}) = \infty$ für jedes feste ρ . Wie Tabelle (3.1) zeigt, liegt der Punkt identischer Effizienz für $T=20$ bei $\rho = .5$.

Von größeren praktischem Interesse ist, aber ein Vergleich der beiden unmittelbar anwendbaren Verfahren OLS und FD. Einerseits gilt hier für festes ρ und $T \rightarrow \infty$ $\text{var}(\hat{\beta}_2^{\text{FD}}) / \text{var}(\hat{\beta}_2^{\text{OLS}}) \rightarrow \infty$, andererseits wegen CHIPMAN (1979, S. 120, Formel (i))

$$\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{GLS}}) / \text{var}(\hat{\beta}_2^{\text{OLS}}) = 5(T^2+T)/6(T^2+1), \quad (3.16)$$

damit wegen (3.15) auch

$$\lim_{\rho \rightarrow 1} \text{var}(\hat{\beta}_2^{\text{FD}}) / \text{var}(\hat{\beta}_2^{\text{OLS}}) = 5(T^2+T)/6(T^2+1) < 1.$$

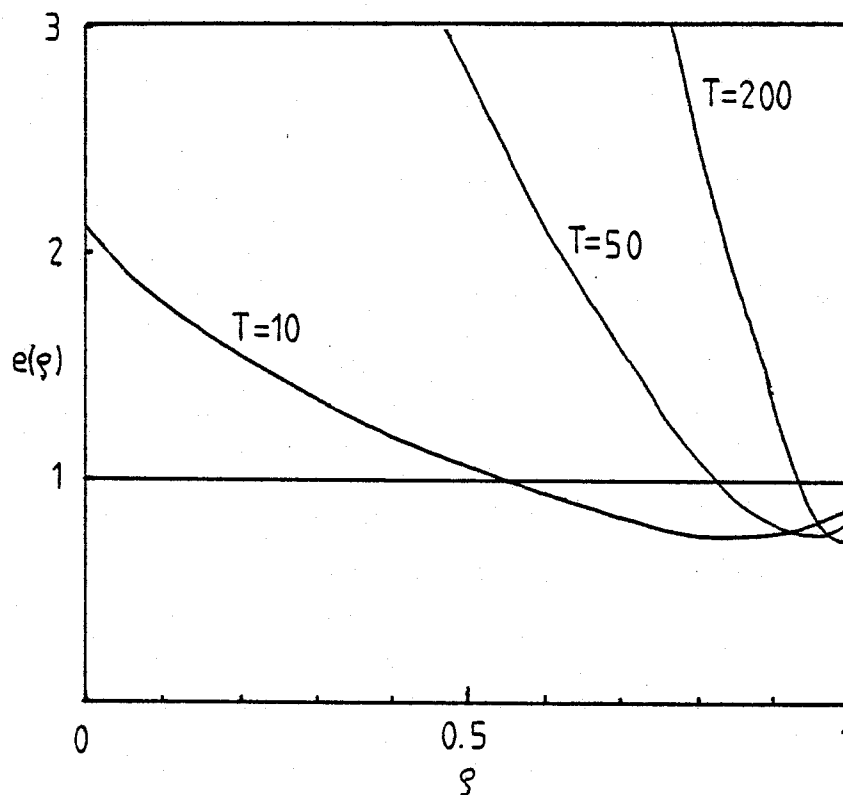
Daher ist FD für beliebiges T für $\rho > \rho^*(T)$ besser als OLS, andererseits OLS bei beliebigem ρ für $T > T^*(\rho)$ besser als FD. Tabelle (3.2) gibt $\rho^*(T)$ für verschiedene T und Schaubild (3.1) einige Kurvenverläufe von $e(\rho) = \text{var}(\hat{\beta}_2^{\text{FD}}) / \text{var}(\hat{\beta}_2^{\text{OLS}})$ wieder (für positives ρ).

TABELLE (3.2): Punkte identischer Effizienz von
OLS und FD für verschiedene T

T	10	20	30	40	50	60	70	80	90
$\rho^*(T)$.572	.726	.799	.841	.869	.888	.902	.913	.922
T	100	200	300	400	500	600	700	800	900
$\rho^*(T)$.929	.963	.975	.981	.985	.987	.989	.990	.991

Wie zunächst die Tabelle zeigt, wird die für $\rho \rightarrow 1$ geltende Überlegenheit von FD für praktisch bedeutsame Stichprobenumfänge erst ab sehr hohen und mit wachsendem T ebenfalls wachsenden Schwellenwerten

SCHAUBILD (3.1): Relative Effizienz von OLS und FD



der Störgrößenkorrelation wirksam. Dieser schließlichen Überlegenheit stehen jedoch erhebliche Effizienzverluste bei mäßigen ρ gegenüber, wie in Schaubild (3.1) (und auch schon in Tabelle (3.1)) zu sehen.

Schaubild (3.1) macht ferner deutlich, daß die relative Effizienz von OLS bei festem T ein relatives Minimum aufweist, und daß die Effizienzkurven $e_T(\rho)$ sich kreuzen. Das folgt aus $e_{T+1}(0) > e_T(0)$ und

$$\lim_{\rho \rightarrow 1} e_{T+1}(\rho) = 5((T+1)^2 + (T+1)) / 6((T+1)^2 + 1) < 5(T^2 + T) / 6(T^2 + 1) = \lim_{\rho \rightarrow 1} e_T(\rho) \quad , \quad (3.17)$$

wobei $\lim_{T \rightarrow \infty} \lim_{\rho \rightarrow 1} e_T(\rho) = 5/6 = .833$.

Schaubild (3.1) dokumentiert außerdem den beträchtlichen Effizienzverlust von FD für kleine ρ . Tabelle (3.3) gibt zur Illustration $e(0)$ und $e(.3)$ für verschiedene T wieder.

TABELLE (3.3): Relative Effizienz von OLS für $\rho = 0, .3$ und verschiedene T

T	10	20	30	40	50	60	70	80	90
$\rho=0$	2.04	3.68	5.35	7.01	8.67	10.34	12.01	13.67	15.34
$\rho=.3$	1.36	2.20	3.08	3.97	4.86	5.76	6.65	7.75	8.44
T	100	200	300	400	500	600	700	800	900
$\rho=0$	17.00	33.67	50.33	67.00	83.67	100.3	117.0	133.7	150.3
$\rho=.3$	9.34	18.31	27.83	36.25	45.23	54.20	63.18	72.15	81.13

Ein letzter Kommentar gilt der Effizienz von OLS relativ zu GLS, für festes ρ und $T \rightarrow \infty$. Wegen stetiger und positiver Spektraldichte des

Störgrößenprozesses lassen sich hier unmittelbar die allgemeinen Resultate von GRENANDER und ROSENBLATT (1957, Kapitel 7) anwenden. Die Regressoren erfüllen (2.8), (2.10) und (2.11) (und sind damit, in der Terminologie von Grenander und Rosenblatt, "langsam wachsend"), und wegen $x_{t1} = 1$, $x_{t2} = t$ ($t=1,2,\dots$) gilt

$$\lim_{T \rightarrow \infty} \frac{\sum_{t=1}^T x_{t+n,i} x_{tj}}{(\sum_{t=1}^T x_{ti})^{1/2} (\sum_{t=1}^T x_{tj})^{1/2}} = 1 \quad (3.18)$$

für $i, j \in \{1,2\}$ und jedes feste $n \in \mathbb{N}$. Damit besteht das Regressorspektrum (siehe GRENANDER und ROSENBLATT, 1957, Abschnitt 7.2) nur aus einem Punkt und es gilt

$$\lim_{T \rightarrow \infty} \frac{\text{var}(\hat{\beta}_i^{\text{OLS}})}{\text{var}(\hat{\beta}_i^{\text{GLS}})} = 1 \quad (i=1,2) . \quad (3.19)$$

Der folgende Unterabschnitt zeigt, daß sich dieses Resultat in gewissem Sinn auch auf trendbehaftete stochastische Regressoren übertragen läßt.

b) bivariate Regression mit exogenen ARIMA Variablen

Das hier betrachtete Modell unterscheidet sich von (3.1) vor allem durch einen ARIMA $(p,1,q)$ -Prozess als unabhängiger Variablen, d.h.

$$x_t - x_{t-1} = \eta_t \quad (t=1,2,\dots) \quad (3.20)$$

mit einem stationären ARMA (p,q) -Prozess $\{\eta_t\}_{t=-\infty}^{\infty}$, $E(\eta_t) = 0$ und

$E(\eta_t^2) = \sigma_\eta^2 > 0$. Dabei soll x_t in diesem Abschnitt allein die t 'te Beobachtung der exogenen Variablen bezeichnen; die sonst so benannte t 'te-Zeile von X ist damit $(1, x_t)$, und die Beobachtungen erfüllen in leichter Verallgemeinerung von (3.1)

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad (t=1,2,\dots) \quad (3.21)$$

Im Gefolge von Box und Jenkins (1970) findet man in der statistischen Literatur zunehmend ARIMA-Prozesse als stochastische Modelle von Zeitreihenbeobachtungen. Ökonometrische Anwendungen umfassen Aktienkurse (siehe etwa GRANGER und MORGENSTERN, 1970; CONRAD und JÜTTNER, 1973; RONNING, 1975, neben vielen anderen), die Umlaufgeschwindigkeit des Geldes (GOULD und NELSON, 1974) oder individuelle Konsumausgaben (HALL, 1978), mit dem Random Walk als ARIMA (0,1,0)-Prozess als wichtigstem Spezialfall. Daher erscheint es sinnvoll, den Konsequenzen von ARIMA-Prozessen als exogenen Variablen in ökonometrischen Modellen nachzugehen.

Im weiteren konzentriert sich das Interesse wie in a) auf Schätzungen des Steigungskoeffizienten β_2 . Daher sei o.B.d.A. (denn die hier betrachteten Schätzverfahren bleiben unter Verschiebungen der x -Achse invariant) $x_0 = 0$, damit

$$E(x_T) = 0, \quad E(x_T^2) = O(T) \quad (T=1,2,\dots) \quad (3.22)$$

und

$$T^{-1} \sum_{t=1}^T x_t^2 \xrightarrow{p} \infty \quad (3.23)$$

Nach Definition (2.21) enthält die exogene Variable damit einen Trend. (3.23) folgt etwa aus HASZA (1980, S. 351, Theorem 1), d.h.

$$T^{-2} \sum_{t=1}^T (x_t - \bar{x})^2 \xrightarrow{D} Z \quad (3.24)$$

mit einer nichtdegenerierten Zufallsvariablen Z . Für spätere Zwecke wird außerdem die Tatsache benötigt, daß Z keine Masse bei 0 besitzt.

In Verallgemeinerung von (3.2) seien die Störgrößen u_t jetzt Ausschnitt eines stationären AR(m)-Prozesses mit beliebigem (aber festen) m , d.h. insbesondere

$$u_t + \rho_1 u_{t-1} + \dots + \rho_m u_{t-m} = \varepsilon_t \quad (t=m+1, m+2, \dots) \quad (3.25)$$

Die ε_t seien $iid(0, \sigma_\varepsilon^2)$, in Verschärfung der Unkorreliertheitsannahme aus a).

Im weiteren werden vor allem die asymptotischen Verteilungen von $\hat{\beta}_2^{\text{OLS}}$ und $\hat{\beta}_2^{\text{GLS}}$ verglichen. Dazu sei folgende Notation vereinbart:

$y = (y_1, \dots, y_T)'$, $u = (u_1, \dots, u_T)'$ (wie in 2a), $1 = (1, 1, \dots, 1)'$,
 $x = (x_1; \dots, x_T)'$ (d.h. $X = (1, x)$), $V = 1/\sigma_\varepsilon^2 E(uu')$. Die Abhängigkeit dieser Größen von T wird auch weiterhin nicht explizit kenntlich gemacht. Ein Punkt über einem Vektor bezeichnet Abweichungen vom jeweiligen Mittelwert (z.B. $\dot{x} = (x_1 - \bar{x}, \dots, x_T - \bar{x})'$). Damit ergibt sich durch Spezialisierung von (2.3)

$$\hat{\beta}_2^{\text{OLS}} = \dot{x}'\dot{y} / \dot{x}'\dot{x} \quad (3.26)$$

Zur Analyse von $\hat{\beta}_2^{\text{GLS}}$ erscheint die Betrachtung von GLS als OLS vorteilhaft, angewandt auf ein transformiertes Modell. Dazu sei R definiert durch $RR' = V$, wobei R so gewählt werden kann, daß R^{-1} untere Dreiecksmatrix ist mit

$$R^{-1} = \begin{matrix} m \\ T-m \end{matrix} \left\{ \begin{array}{cccccccc} x & & & & & & & \\ x & x & & & & & & \\ \vdots & & \ddots & & & & & \\ x & \dots & \dots & \dots & x & & & \\ \rho_m & \rho_{m-1} & \dots & \rho_1 & 1 & & & \\ 0 & \rho_m & \rho_{m-1} & \dots & \rho_1 & 1 & & \\ \vdots & & & & & & \ddots & \\ 0 & \dots & 0 & \rho_m & \rho_{m-1} & \dots & \rho_1 & 1 \end{array} \right\} \quad (3.27)$$

Ferner sei $\tilde{y} = R^{-1}y$, $\tilde{1} = R^{-1}1$, $\tilde{x} = R^{-1}x$, $\tilde{X} = R^{-1}X = (\tilde{1}, \tilde{x})$, und $\tilde{u} = R^{-1}u$. Mit diesen Vereinbarungen ist $\tilde{y} = \tilde{X}\beta + \tilde{u}$ und $\hat{\beta}_2^{\text{GLS}}$ das zweite Element in

$$\begin{pmatrix} \tilde{1}'\tilde{1} & \tilde{1}'\tilde{x} \\ \tilde{x}'\tilde{1} & \tilde{x}'\tilde{x} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{1}'\tilde{y} \\ \tilde{x}'\tilde{y} \end{pmatrix}. \quad (3.28)$$

Die Herleitung der asymptotischen Verteilung von $\hat{\beta}_2^{\text{GLS}}$ erfordert einige vorbereitende Lemmata. Dazu sei $\tilde{\dot{x}} = R^{-1}\dot{x} = R^{-1}(I-D)x$ mit $D = (d_{ij})$ und $d_{ij} = 1/T$ ($i, j=1, \dots, T$) (während $\dot{\tilde{x}} = (I-D)R^{-1}x \neq \tilde{\dot{x}}$ wegen $DR^{-1} \neq R^{-1}D$). Analog sei $\tilde{\dot{u}} = R^{-1}(I-D)u$ und $\dot{\tilde{u}} = (I-D)R^{-1}u$.

LEMMA (3.2):

$$T^2 / \tilde{\dot{x}}'\tilde{\dot{x}} = o_p(1) \quad (3.29)$$

BEWEIS: Sei $\theta_{\max}(T)$ der maximale Eigenwert von V . Wegen der AR(m)-Eigenschaft der u_t 's gilt $\theta_{\max}(T) = O(1)$, damit aber auch

$$T^2 / \tilde{x}' \tilde{x} = T^2 / \dot{x} V^{-1} \dot{x} \leq \theta_{\max}(T) T^2 / \dot{x}' \dot{x} = O_p(1),$$

denn wegen (3.24) gilt $T^2 / \dot{x}' \dot{x} = O_p(1)$, q.e.d.

Eine äquivalente und im weiteren alternativ zu (3.29) gebrauchte Schreibweise für Lemma (3.2) ist

$$T / \|\tilde{x}\| = O_p(1).$$

Weiterhin folgt aus (3.29) sofort, daß $\tilde{x}' \tilde{x} = O_p(T^2)$, aber die Formulierung (3.29) betont, daß $\tilde{x}' \tilde{x} / T^2 \xrightarrow{p} 0$, eine durch die Gleichung $\tilde{x}' \tilde{x} = O_p(T^2)$ nicht ausgeschlossene Möglichkeit (dabei wird die $o, 0$ -Notation in der allgemein üblichen, etwa bei FULLER (1976, Kap. 5) nachzuschlagenden Bedeutung gebraucht).

LEMMA (3.3):

$$\tilde{x}' \tilde{x} / \dot{x}' \dot{x} \xrightarrow{p} 1. \quad (3.30)$$

BEWEIS: Wegen $\dot{x} = (I-D)R^{-1}x$ und $\tilde{x} = R^{-1}(I-D)x$ gilt $\tilde{x} - \dot{x} = (DR^{-1} - R^{-1}D)x$, wobei die rechte untere $(T-m) \times (T-m)$ -Teilmatrix von $DR^{-1} - R^{-1}D$ verschwindet. Daher gilt $\|\tilde{x} - \dot{x}\| = O_p(T^{1/2})$, und damit

$$\|\tilde{x}\| / \|\dot{x}\| \leq 1 + \|\tilde{x} - \dot{x}\| / \|\dot{x}\| \xrightarrow{p} 1. \quad (3.31)$$

Analog folgt durch Vertauschung der Rollen von \tilde{x} und \dot{x} , daß $\text{plim} \|\dot{x}\| / \|\tilde{x}\| \geq 1$, und damit (3.30). q.e.d.

LEMMA (3.4):

$$(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{1/2}(\hat{\beta}_2^{\text{GLS}} - \beta_2) = \tilde{\mathbf{x}}'\tilde{\mathbf{u}} / (\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{1/2} + o_p(1) \quad (3.32)$$

BEWEIS: Aus (3.28) folgt nach einigen Umformungen

$$\hat{\beta}_2^{\text{GLS}} - \beta_2 = \dot{\tilde{\mathbf{x}}}'\dot{\tilde{\mathbf{u}}} / \dot{\tilde{\mathbf{x}}}'\dot{\tilde{\mathbf{x}}} + o_p(T^{-1}), \quad (3.33)$$

d.h. das Analogon zu (3.26). Da die erste Spalte des transformierten Modells nicht mehr aus Konstanten besteht (erst ab Zeile m gilt $\tilde{1}_t = \sum_{i=0}^m \rho_i \neq 0$), gibt es keine exakte Entsprechung. Lemma (3.2) und Lemma (3.3) liefern die Ordnung des Resttermes, und wegen

$$\dot{\tilde{\mathbf{x}}}'\dot{\tilde{\mathbf{u}}} = \dot{\tilde{\mathbf{x}}}'\tilde{\mathbf{u}} \quad \text{und} \quad \|\dot{\tilde{\mathbf{x}}}'\tilde{\mathbf{u}} - \tilde{\mathbf{x}}'\tilde{\mathbf{u}}\| = o_p(T^{1/2})$$

auch (3.32) .

q.e.d.

Die in (3.32) gewählte Normierung erweist sich als nützlich sowohl zur Herleitung der asymptotischen Verteilung von $\hat{\beta}_2^{\text{GLS}}$ als auch zum Vergleich mit $\hat{\beta}_2^{\text{OLS}}$.

SATZ (3.5): Unter den allgemeinen Voraussetzungen von Kapitel 2 und den Zusatzbedingungen dieses Abschnitts gilt

$$(\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{1/2}(\hat{\beta}_2^{\text{GLS}} - \beta_2) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2) . \quad (3.34)$$

BEWEIS: Wegen Lemma (3.4) ist die Grenzverteilung in (3.34) identisch mit derjenigen der standardisierten OLS-Schätzung im univariaten Modell

$$y = \tilde{x} \beta + \tilde{u} .$$

Wegen (3.27) gilt ferner für $t > m$

$$\tilde{u}_t = u_t + \rho_1 u_{t-1} + \dots + \rho_m u_{t-m} = \varepsilon_t , \quad (3.35)$$

d.h. die Störgrößen erfüllen bis auf eine vernachlässigbare Menge zu Beginn der Zeitreihe die Bedingungen von Satz (2.3). Zur unmittelbaren Anwendung dieses Satzes bleibt allein (2.22) nachzuweisen, im vorliegenden Spezialfall

$$\max_{1 \leq t \leq T} \tilde{x}_t^2 / \tilde{x}' \tilde{x} \xrightarrow{P} 0 . \quad (3.36)$$

Zum Beweis von (3.36) beachte man, daß wegen (3.27) und in Analogie zu (3.35)

$$\tilde{x}_t = \dot{x}_t + \rho_1 \dot{x}_{t-1} + \dots + \rho_m \dot{x}_{t-m} \quad (3.37)$$

für $t > m$. Wegen $\text{var}(\dot{x}_T) = O(T)$ gilt damit auch $\text{var}(\tilde{x}_T) = O(T)$ und $\tilde{x}_T / T \xrightarrow{P} 0$. Dies liefert wegen Lemma (3.2) aber sofort

$$\tilde{x}_T^2 / \tilde{x}' \tilde{x} \xrightarrow{P} 0 . \quad (3.38)$$

Nun sei $t^*(T)$ ($1 \leq t^*(T) \leq T$) der Index, für den $\tilde{x}_t^2 / \tilde{x}' \tilde{x}$ das Maximum annimmt. $\{t^*(T)\}_{T=1,2,\dots}$ ist eine Folge von Zufallsvariablen mit Werten in \mathbb{N} , sowie $t^*(T) \leq t^*(T+1)$ ($T=1,2,\dots$) und

$$t^*(T) \xrightarrow{P} \infty . \quad (3.39)$$

Daher gilt

$$\begin{aligned} \max_{1 \leq t \leq T} \frac{\tilde{x}_t^2}{\tilde{x}'\tilde{x}} &= \frac{\tilde{x}_{t^*(T)}^2}{\sum_{t=1}^T \tilde{x}_t^2} \\ &\leq \frac{\tilde{x}_{t^*(T)}^2}{\sum_{t=1}^{t^*(T)} \tilde{x}_t^2} \\ &\xrightarrow{P} 0 \quad \text{q.e.d.} \end{aligned} \quad (3.40)$$

Wegen Lemma (3.2) liefert Satz (3.5) sofort $\hat{\beta}_2^{\text{GLS}} - \beta_2 = o_p(T^{-1})$, im Vergleich zu $o_p(T^{-1/2})$ im Standardfall ohne Trend. Insbesondere ist GLS damit (schwach) konsistent. Die Konvergenz gegen Null des Schätzfehlers ist aber nicht so schnell wie im Modell mit deterministischem Trend aus Abschnitt a). Da $\tilde{x}'\tilde{x}/T^2$ nicht notwendig gegen eine Konstante konvergiert (siehe Hasza, 1980), läßt sich Satz (3.5) auch nicht auf die übliche Weise in eine Aussage zur Grenzverteilung von $T(\hat{\beta}_2^{\text{GLS}} - \beta_2)$ übersetzen.

Zur Konstruktion von (asymptotischen) Signifikanztests und Konfidenzintervallen für β_2 bleibt noch die Konsistenz der GLS-Schätzung

$$s^2 = (\tilde{y} - \tilde{X}\hat{\beta}^{\text{GLS}})'(\tilde{y} - \tilde{X}\hat{\beta}^{\text{GLS}})/(T-2)$$

für σ_ε^2 zu zeigen. Dies folgt aber unmittelbar aus Satz (2.4), angewandt auf das transformierte Modell $\tilde{y} = \tilde{X}\beta + \tilde{u}$. Damit gilt

$$\frac{(\tilde{x}'\tilde{x})^{1/2}}{s} (\hat{\beta}_2^{\text{GLS}} - \beta_2) \xrightarrow{D} N(0,1) \quad (3.41)$$

Das nächst Resultat gilt dem Vergleich von OLS und GLS und basiert auf folgendem fundamentalem Lemma:

LEMMA (3.6): Unter den Voraussetzungen dieses Abschnitts gilt für alle $m, n \in \mathbb{N}$ mit $m > n$

$$\frac{\sum_{t=m}^T \dot{x}_t \dot{x}_{t-n}}{\sum_{t=1}^T \dot{x}_t^2} \xrightarrow{P} 1 \quad (3.42)$$

BEWEIS: HASZA(1980, S. 351, theorem 2) zeigt, daß

$$T(1 - \frac{\sum_{t=m}^T \dot{x}_t \dot{x}_{t-n}}{\sum_{t=1}^T \dot{x}_t^2}) \xrightarrow{D} V_n \quad (3.43)$$

mit einer endlichen und wohldefinierten Zufallsvariablen V_n . Daraus folgt

$$1 - \frac{\sum_{t=m}^T \dot{x}_t \dot{x}_{t-n}}{\sum_{t=1}^T \dot{x}_t^2} \xrightarrow{P} 0,$$

in Äquivalenz zu (3.42).

q.e.d.

SATZ (3.7): Unter den allgemeinen Voraussetzungen von Kapitel 2 und den Zusatzbedingungen dieses Abschnitts sind die Grenzverteilungen von $(\tilde{x}'\tilde{x})^{1/2}(\hat{\beta}_2^{\text{GLS}} - \beta_2)$ und $(\tilde{x}'\tilde{x})^{1/2}(\hat{\beta}_2^{\text{OLS}} - \beta_2)$ identisch.

BEWEIS: Der Satz folgt unmittelbar aus den folgenden im weiteren zu beweisenden Hilfsresultaten

$$(\tilde{x}'\tilde{x})^{1/2}(\hat{\beta}_2^{\text{OLS}} - \beta_2) = (\tilde{x}'\tilde{x})^{-1/2} \tilde{x}'\tilde{u} (\sum_{i=0}^m \rho_i)^2 + o_p(1) \quad (3.44)$$

und

$$(\tilde{x}'\tilde{x})^{1/2}(\hat{\beta}_2^{\text{GLS}} - \beta_2) = (\tilde{x}'\tilde{x})^{-1/2} \tilde{x}'\tilde{u} (\sum_{i=0}^m \rho_i)^2 + o_p(1) \quad (3.45)$$

Zum Beweis von (3.44) beachte man, daß wegen (3.26)

$$(\tilde{\mathbf{x}}' \tilde{\mathbf{x}})^{1/2} (\hat{\beta}_2^{\text{OLS}} - \beta_2) = \frac{\tilde{\mathbf{x}}' \tilde{\mathbf{u}}}{(\tilde{\mathbf{x}}' \tilde{\mathbf{x}})^{1/2}} \cdot \frac{\tilde{\mathbf{x}}' \tilde{\mathbf{x}}}{\tilde{\mathbf{x}}' \tilde{\mathbf{x}}} \quad (3.46)$$

gilt, wobei der erste Quotient stochastisch beschränkt bleibt und

$$\tilde{\mathbf{x}}' \tilde{\mathbf{x}} / \mathbf{x}' \mathbf{x} = \left(\sum_{i=0}^m \rho_i \right)^2 + o_p(1) \quad (3.47)$$

Der kritische Punkt ist der Beweis von (3.47), woraus (3.44) wegen (3.46) unmittelbar folgt. Dazu beachte man, daß für $t > m$, $\tilde{\mathbf{x}}_t = \sum_{i=0}^m \rho_i \mathbf{x}_{t-i}$ und daher

$$\tilde{\mathbf{x}}' \tilde{\mathbf{x}} = \sum_{t=1}^T \tilde{\mathbf{x}}_t^2 = \sum_{t=m+1}^T \left(\sum_{i=0}^m \rho_i \mathbf{x}_{t-i} \right)^2 + o_p(T) \quad .$$

Daraus folgt aber wegen Lemma (3.6)

$$\begin{aligned} \tilde{\mathbf{x}}' \tilde{\mathbf{x}} / \mathbf{x}' \mathbf{x} &= \sum_{i,j=0}^m \rho_i \rho_j \left(\sum_{t=m+1}^T \mathbf{x}_{t-i} \mathbf{x}_{t-j} \right) / \sum_{t=1}^T \mathbf{x}_t^2 + o_p(1) \\ &= \left(\sum_{i=0}^m \rho_i \right)^2 + o_p(1) \quad , \end{aligned}$$

als Beweis von (3.47) und damit auch (3.44).

Zum Beweis von (3.45) reicht es wegen (3.32) und (3.29) zu zeigen, daß

$$\tilde{\mathbf{x}}' \tilde{\mathbf{u}} = \left(\sum_{i=0}^m \rho_i \right)^2 \mathbf{x}' \tilde{\mathbf{u}} + o_p(T) \quad (3.48)$$

Dies folgt aber aus

$$\begin{aligned} \tilde{\mathbf{x}}' \tilde{\mathbf{u}} &= \sum_{t=m+1}^T \left(\sum_{i=0}^m \rho_i \mathbf{x}_{t-i} \right) \left(\sum_{j=0}^m \rho_j \mathbf{u}_{t-j} \right) + o_p(1) \\ &= \sum_{i,j=0}^m \rho_i \rho_j \sum_{t=m+1}^T \mathbf{x}_{t-i} \mathbf{u}_{t-j} + o_p(1) \\ &= \sum_{i,j=0}^m \rho_i \rho_j \sum_{t=1}^T \mathbf{x}_t \mathbf{u}_t + o_p(T^{1/2}) \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{i=0}^m \rho_i \right)^2 \dot{x}'u + o_p(T) \\
&= \left(\sum_{i=0}^m \rho_i \right)^2 \dot{x}'\dot{u} + o_p(T) .
\end{aligned}$$

Dabei ist das erste Gleichheitszeichen gerechtfertigt wegen

$$\sum_{t=m+1}^T \dot{x}_{t-i} u_{t-j} = \sum_{t=m+1}^T \dot{x}_{t-j} u_{t-j} + \sum_{t=m+1}^T (\dot{x}_{t-i} - \dot{x}_{t-j}) u_{t-j} ,$$

wobei $\text{var}(\sum_{t=m+1}^T (\dot{x}_{t-i} - \dot{x}_{t-j}) u_{t-j}) = O(T)$,

d.h. $\sum_{t=m+1}^T (\dot{x}_{t-i} - \dot{x}_{t-j}) u_{t-j} = o_p(T^{1/2})$.

Daraus folgt auch

$$\sum_{i,j=0}^m \rho_i \rho_j \sum_{t=m+1}^T \dot{x}_{t-i} u_{t-j} = \sum_{i,j=0}^m \rho_i \rho_j \sum_{t=1}^T \dot{x}_t u_t + o_p(T^{1/2}) .$$

Die letzte Gleichung zum Beweis von (3.48) schließlich folgt aus $\dot{x}'u = \dot{x}'\dot{u}$, was den Beweis von (3.45) und damit von Satz (3.7) komplettiert. q.e.d.

Satz (3.7) zeigt damit die asymptotische Äquivalenz von OLS und GLS auch unter den Annahmen von Abschnitt b). Wie in Abschnitt a) bleiben hier "feasible" GLS-Versionen außer Betracht. Jedoch sieht man sofort, daß auch im gegenwärtigen Modell die FD-Schätzung für β_2 , die für hohe Störgrößenkorrelation oft als Approximation für GLS herangezogen wird, eine asymptotisch verschwindende Effizienz besitzt (für beliebige, aber feste ρ_i und $T \rightarrow \infty$). Wegen $\bar{D}x'Dx = \sum_{t=1}^T \eta_t^2 = o_p(T)$ gilt $\hat{\beta}_2^{FD} - \beta_2 = o_p(T^{-1/2})$, verglichen mit $o_p(T^{-1})$ für OLS und GLS.

c) Simulation

Dieser Abschnitt überprüft mit Hilfe von Monte-Carlo Simulationen die Aussagekraft der asymptotischen Resultate aus b) für endliche Stichproben. Dabei werden durch Spezialisierung von (3.20) die Zeitreihen der exogenen Variablen durchweg als Random Walk erzeugt, d.h.

$x_t = x_{t-1} + \eta_t$ mit unabhängigen und identisch verteilten η_t ($t=1,2,\dots$). Dies scheint zugleich der für ökonometrische Anwendungen wichtigste ARIMA-Prozess zu sein. Die Störgrößen u_t werden wie in Abschnitt a) als AR(1)-Prozess modelliert, ebenfalls der in ökonometrischen Anwendungen bei weitem häufigste Spezialfall.

Alle Experimente wurden mit eigenen Programmen auf dem Sperry-Univac 1100 -Rechner des Instituts für Höhere Studien durchgeführt. Zufallszahlen wurden mit den NAG-Routinen G05CAF und G05DDF erzeugt (siehe NUMERICAL ALGORITHMS GROUP, 1981, Library Manual, Volume 6).

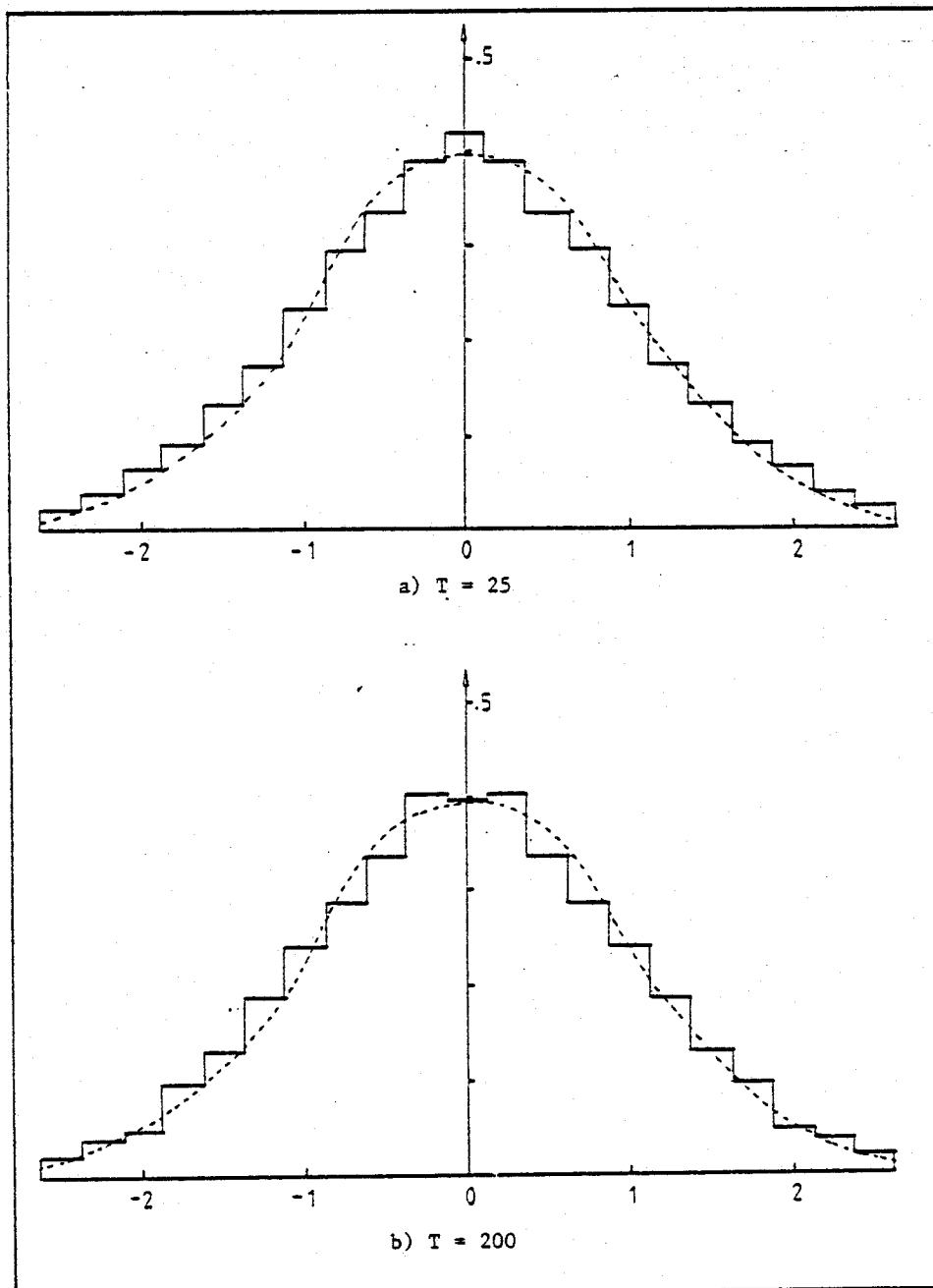
Eine erste Serie von Experimenten gilt der endlichen-Stichproben-Relevanz von Satz (3.5), d.h. der Güte der Normalverteilungsapproximation. Dabei wird hier nur der Fall unabhängiger Störgrößen, d.h. OLS = GLS untersucht. Das erscheint deswegen sinnvoll, weil für praktische Anwendungen vor allem OLS und weniger das nur theoretisch bedeutsame GLS interessiert.

Schaubild (3.2) gibt die Histogramme von 10 000 Realisationen von

$$\frac{(\dot{x}'\dot{x})^{1/2}}{s} (\hat{\beta}_2^{\text{OLS}} - \beta_2) \quad (3.49)$$

an, für $T=25$ (3.2-a) und $T=200$ (3.2-b). Wegen Satz (3.5) und (3.41) ist (3.49) asymptotisch standardnormalverteilt, unabhängig von der Verteilung der η_t 's und u_t 's, sofern diese ansonsten die Voraussetzungen dieses

SCHAUBILD (3.2): Histogramm von 10 000 Realisationen von OLS
(Störgrößen normalverteilt)



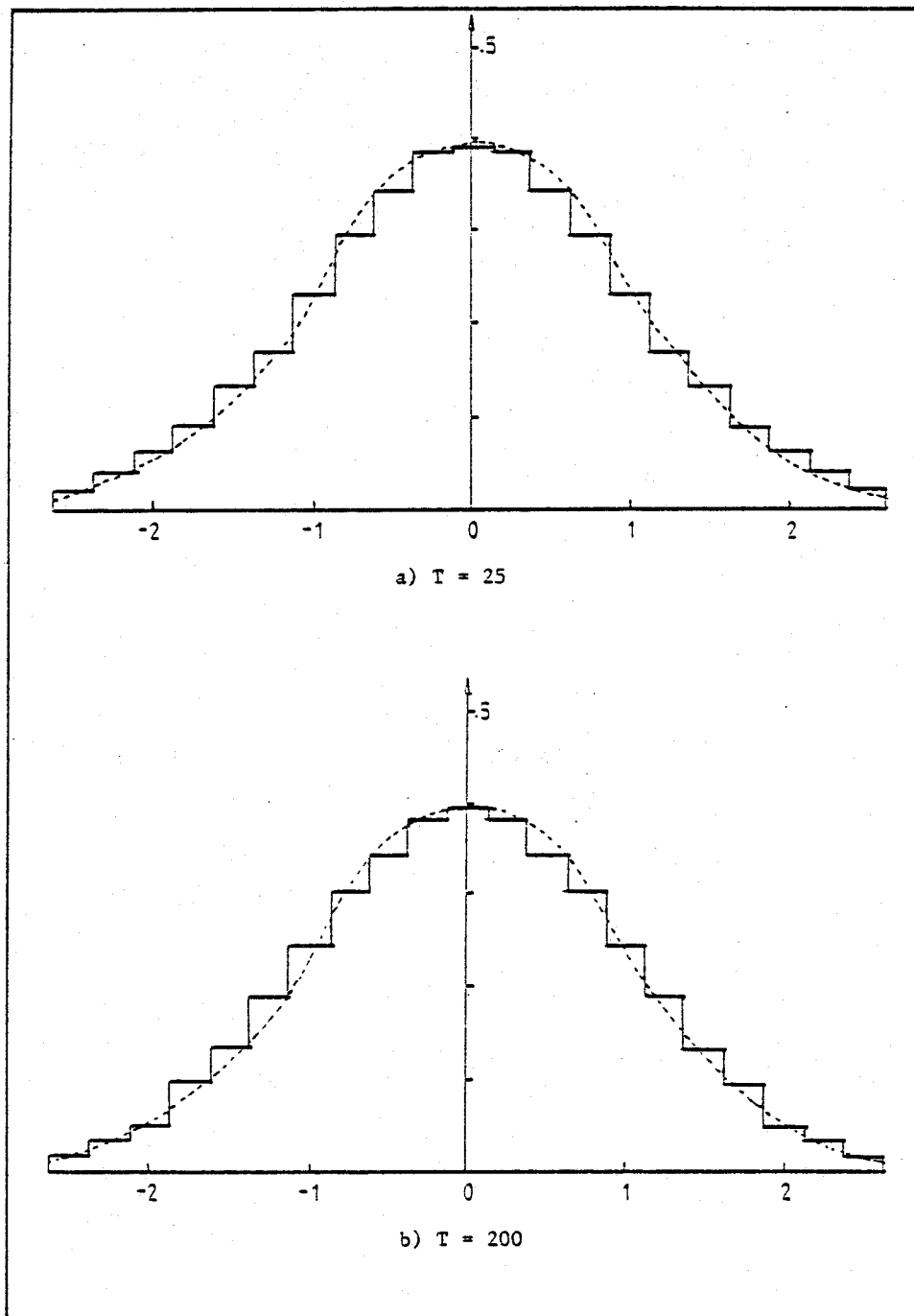
Kapitels erfüllen. Zum Vergleich ist auch die Dichtefunktion der Normalverteilung eingezeichnet.

Für Schaubild (3.2) sind sowohl die η_t 's als auch die u_t 's selbst standardnormalverteilt. Weiter wurde in allen Experimenten $\beta_1 = \beta_2 = 1$ gesetzt. Das dient allein zur Vereinfachung der Rechnungen, denn wahre Verteilung und Monte-Carlo Simulation von $\hat{\beta}_2^{\text{OLS}} - \beta_2$ hängen von diesen Parametern nicht ab. Analog ist die Verteilung von (3.49) unabhängig von $\sigma_\eta^2 = \text{var}(\eta_t)$ und $\sigma_u^2 = \text{var}(u_t)$, so daß die hier gewählten Werte für β_1 , β_2 , $\sigma_u^2 (=1)$ und $\sigma_\eta^2 (=1)$ auch Rückschlüsse auf andere Parameterkombinationen erlauben. Der einzige, die wahre und empirische Verteilung von (3.49) beeinflussende Parameter ist der Stichprobenumfang T .

Die Intervallbreite der Histogramme ist .25 , von $z = -2.625$ bis $z = 2.625$. Wegen symmetrischer Verteilung von $\hat{\beta}_2^{\text{OLS}} - \beta_2$ wird nach einer Anregung von DICKEY und FULLER (1981) die Höhe des Histogramms als arithmetische Mittel je zweier symmetrischer Intervalle berechnet. Wegen negativer Korrelation der Besetzungszahlen je zweier Intervalle (ob symmetrisch oder nicht) liefert das eine sonst nur mit mehr als 20 000 Wiederholungen zu erzielende Genauigkeit der Monte-Carlo Schätzungen der Besetzungswahrscheinlichkeiten. Weitere "sampling tricks" wurden nicht verwendet .

Wie das Schaubild zeigt, scheint schon für $T=25$ die asymptotische Normalverteilung die wahre Verteilung von (3.49) gut zu approximieren. Um zu überprüfen, inwieweit das auf die Normalverteilung der u_t 's selbst zurückgeht, wurden in einer weiteren Versuchsreihe die Störgrößen als gleichverteilte Zufallszahlen erzeugt. Schaubild (3.3) gibt die zu Schaubild (3.2) analogen Ergebnisse wieder. Auch hier scheint schon für $T=25$ die Normalverteilung eine zufriedenstellende Approximation zu liefern.

SCHAUBILD (3.3): Histogramm von 10 000 Realisationen von OLS
(Störgrößen gleichverteilt)



Zur Präzisierung von "zufriedenstellende Approximation" gibt Tabelle (3.5) Monte-Carlo Schätzungen der Verteilungsfunktion von (3.49) für verschiedene T und Ordinatenwerte an. Zum Vergleich sind wiederum die entsprechenden Werte für die Standardnormalverteilung beigelegt. Die letzte Zeile der Tabelle enthält das Maximum der geschätzten Standardabweichungen der Monte-Carlo Ergebnisse der jeweiligen Spalte.

Wie schon in den Histogrammen wurden die Monte-Carlo Schätzungen der einzelnen Wahrscheinlichkeiten durch Mittelung der Simulationsergebnisse jeweils symmetrischer Halbintervalle gewonnen. Das ergäbe für die letzte Spalte durchweg den Wert .5 , den (wegen Symmetrie der Verteilung bekannten) wahren Wert der Wahrscheinlichkeit. Daher gibt die letzte Spalte zu Kontrollzwecken allein den Anteil der negativen Ausprägungen von (3.39) wieder.

TABELLE (3.4): Vergleich von relativen Häufigkeiten und Grenzwahrscheinlichkeiten (Störgrößen gleichverteilt)

	Wahrscheinlichkeit/rel. Häufigkeit, daß ZV kleiner/gleich				
	-2.625	-2	-1	-.5	0
T=25	.0082	.0299	.1617	.3113	.4977
T=50	.0064	.0255	.1610	.3092	.5001
T=100	.0054	.0246	.1614	.3077	.5018
T=200	.0041	.0219	.1619	.3122	.4990
N(0,1)	.0044	.0228	.1587	.3085	.5000
Stichproben- fehler	.0006	.0011	.0025	.0032	.0050

Die Tabelle deutet darauf hin, daß die Wahrscheinlichkeiten absolut großer Werte von (3.49) durch die Normalverteilungsapproximation unterschätzt werden. Ein zweiseitiger Signifikanztest für $\beta_2 = 0$ mit nominalem Signifikanzniveau $\alpha = 4.56\%$ beinhaltet damit bei $T = 25$ eine wahre Wahrscheinlichkeit für einen Fehler erster Art von etwa 6% . Ähnliche Ergebnisse erbrachten auch Simulationen mit normalverteilten Störgrößen.

Eine weitere Serie von Experimenten überprüft die endliche-Stichproben Relevanz von Satz (3.7), der (im dort spezifizierten Sinn) asymptotischen Äquivalenz von OLS und GLS. Dazu werden die empirischen Verteilungen von OLS und GLS verglichen an Hand von (i) Histogrammen der Monte-Carlo Ergebnisse wie in den Schaubildern (3.2) und (3.3), (ii) den relativen Häufigkeiten von $|\tilde{x}'\tilde{x}^{-1/2}/s (\hat{\beta}_2^{\text{GLS/OLS}} - \beta_2)| > z$ für verschiedene z , und (iii) durch die empirischen Varianzen von $\hat{\beta}_2^{\text{GLS}}$, $\hat{\beta}_2^{\text{OLS}}$ und $\hat{\beta}_2^{\text{OLS}} - \hat{\beta}_2^{\text{GLS}}$.

BASMAN (1961) kritisiert empirische Momente zur Beschreibung von Wahrscheinlichkeitsverteilungen bei Monte-Carlo Experimenten mit dem Hinweis, daß bei Nichtexistenz der entsprechenden Populationsparameter mit wachsender Versuchsanzahl keine Konvergenz eintritt. Das trifft jedoch bei computergestützten Monte-Carlo Experimenten, also dem bei weitem häufigsten Regelfall, nicht zu. Da jede Rechenanlage nur Zahlen bis zu einem von der jeweiligen Hardware abhängenden Maximum z verarbeiten kann, geschieht (bei Vernachlässigung sonstiger numerischer Probleme) jede computergestützte Monte-Carlo Simulation einer Zufallsvariablen Y nur für die bedingte Verteilung von Y , gegeben $|Y| \leq z$. Diese so gestutzte Verteilung hat aber Momente jeder Ordnung.

Außer von T hängen die zu simulierenden Größen nur noch Autoregressionsparameter ρ der Störgrößen ab (nicht aber von β_1 und β_2 , die weiterhin =1 gesetzt werden). Auch σ_η^2 und σ_u^2 behalten weiter den Wert 1. Das bedeutet keine Einschränkung der Allgemeinheit, da die im folgenden vor allem betrachteten Quotienten empirischer Varianzen von den konkret bei der Simulation gewählten Werten von σ_η^2 und σ_u^2 nicht abhängen.

Für jede Kombination von T und ρ werden 200 Schätzwerte $\hat{\beta}_2^{\text{GLS}}$ und $\hat{\beta}_2^{\text{OLS}}$ erzeugt, jeweils mit neuen Zufallszahlen. Daraus werden die empirischen Varianzen von $\hat{\beta}_2^{\text{GLS}}$ ($=\text{var}(\text{GLS})(e)$), $\hat{\beta}_2^{\text{OLS}}$ ($=\text{var}(\text{OLS})(e)$) und $\hat{\beta}_2^{\text{OLS}} - \hat{\beta}_2^{\text{GLS}}$ ($=\text{var}(\text{OLS-GLS})(e)$) ermittelt. Im weiteren wird allein

$$1 + \text{var}(\text{OLS-GLS})(e) / \text{var}(\text{GLS})(e) \quad (3.50)$$

explizit wiedergegeben. Falls $\text{var}(\hat{\beta}_2^{\text{OLS}})$ und $\text{var}(\hat{\beta}_2^{\text{GLS}})$ existieren, ist (3.50) eine effizientere Schätzung für

$$E\{(\hat{\beta}_2^{\text{OLS}} - \beta_2)^2\} / E\{(\hat{\beta}_2^{\text{GLS}} - \beta_2)^2\} \quad (3.51)$$

als $\text{var}(\text{OLS})(e) / \text{var}(\text{GLS})(e)$, da (3.50) die Kenntnis verwendet, daß $\text{cov}(\hat{\beta}_2^{\text{OLS}} - \hat{\beta}_2^{\text{GLS}}, \hat{\beta}_2^{\text{GLS}}) = 0$. Außerdem erlaubt (3.50) einen unmittelbaren Rückschluß auf die relative Abweichung von OLS und GLS, verglichen mit dem Schätzfehler von GLS.

Existenz vorausgesetzt, sind Zähler und Nenner von (3.50) grundsätzlich auch analytisch berechenbar. Man hat

$$\begin{aligned} \text{cov}(\hat{\beta}^{\text{OLS}}) &= \sigma_\varepsilon^2 E\{(X'X)^{-1} X'VX(X'X)^{-1}\} , \\ \text{cov}(\hat{\beta}^{\text{GLS}}) &= \sigma_\varepsilon^2 E\{(X'V^{-1}X)^{-1}\} , \end{aligned} \quad (3.52)$$

aber für das gegenwärtige beschränkte Untersuchungsziel lassen die

Vorteile einer Monte-Carlo Simulation als "capital intensive approach" (SUMMERS, 1965) eine mühsame analytische Ableitung (den "labour intensive approach") wenig sinnvoll erscheinen.

Tabelle (3.5) gibt Simulationsergebnisse für normalverteilte η_t und u_t wieder, für verschiedene T und ρ . Als Hauptergebnis zeigt sich, daß die Annäherung der Verteilungen von OLS und GLS (gemessen durch $\text{var}(\text{OLS}/\text{GLS})(e)$ keineswegs monoton in T erfolgt. Zwar ist für kleine ρ schon bei bescheidenen Stichprobenumfängen gute Übereinstimmung festzustellen, aber für große positive ρ nimmt der Quotient der empirischen Varianzen und die relative Abweichung von OLS und GLS verglichen mit dem Schätzfehler von GLS zunächst zu.

Eine mögliche Erklärung für dieses Phänomen liefert CHIPMAN'S (1979) Analyse des in a) untersuchten verwandten Modells mit deterministischem Trend. Dort (S. 124, Tabelle II) wird gezeigt, daß der Punkt minimaler relativer Effizienz von OLS mit wachsendem T monoton gegen 1 strebt. Gleichzeitig nimmt das Minimum selbst monoton ab (d.h. die Konvergenz gegen 1 der relativen Effizienz von OLS für $T \rightarrow \infty$ ist nicht gleichmäßig für $\rho \in (-1,1)$). Damit nimmt für ein gegebenes ρ , das für kleines T

TABELLE (3.5): Quotienten der empirischen Varianzen
(200 Wiederholungen; Störgrößen normalverteilt)

T	ρ							
	-0.95	-0.9	-0.6	-0.3	.3	.6	.9	.95
25	6.07	3.56	1.45	1.07	1.09	1.50	4.56	8.25
50	5.17	3.22	1.18	1.05	1.10	1.62	8.16	10.07
100	2.72	1.63	1.18	1.02	1.03	1.26	4.59	10.88
200	1.76	1.31	1.07	1.01	1.02	1.17	4.47	9.30
400	1.51	1.18	1.03	1.01	1.01	1.12	2.90	9.64

rechts vom Minimum liegt, die relative Effizienz von OLS mit wachsendem T zunächst ab. Tabelle (3.5) deutet darauf hin, daß ähnliches auch im vorliegenden Modellzusammenhang gilt.

Tabelle (3.6) liefert die analogen Zahlen für gleichverteilt ε_t (die Störgrößen u_t selbst sind dann nicht gleichverteilt). Das dient vor allem der Kontrolle der Variabilität der Monte-Carlo Ergebnisse, denn Zähler und Nenner von (3.51) hängen, Existenz vorausgesetzt, von der konkreten Verteilung der Störgrößen nicht ab.

TABELLE (3.6): Quotienten der empirischen Varianzen
(200 Wiederholungen; Störgrößen gleichverteilt)

T	ρ							
	-.95	-.9	-.6	-.3	.3	.6	.9	.95
25	5.82	3.39	1.44	1.10	1.10	1.50	5.73	6.90
50	4.03	2.46	1.20	1.03	1.10	1.42	4.85	10.24
100	2.46	1.88	1.11	1.04	1.05	1.29	6.95	11.39
200	2.55	1.60	1.08	1.01	1.04	1.23	3.91	11.29
400	1.63	1.16	1.03	1.01	1.01	1.09	2.87	8.04

Trotz geringfügiger Abweichungen, vor allem für große ρ , bestätigt Tabelle (3.6) die früheren Ergebnisse. Für absolut große ρ und kleine T ist OLS trotz Satz (3.7) sehr ineffizient, und diese Ineffizienz nimmt sogar für hohe positive ρ mit wachsendem T zunächst noch zu. Auch alternative Effizienzmaße führen zu keiner anderen Aussage. Damit bleibt als Hauptergebnis dieses Abschnitts festzuhalten, daß für bestimmte Parameterkonstellationen die praktische Relevanz von Satz (3.7) erst bei sehr großen Stichproben einsetzt.

d) Diskussion

Ein bemerkenswertes Teilergebnis dieses Kapitels ist die für $T \rightarrow \infty$ verschwindende relative Effizienz des Erste-Differenzen Schätzers FD. Auf diese Konsequenz von Trend in den hier betrachteten Modellen weist schon HANNAN (1960, S. 115) hin. In scharfem Gegensatz dazu steht die bei gegebenem T gegen Eins strebende Effizienz von FD relativ zu GLS für den Fall wachsender Störgrößenkorrelation. Das folgt für das einfache Modell aus Abschnitt a) aus (3.15) und wird in Anhang b für beliebige Regressionsmodelle mit festen Regressoren und AR-Störgrößen gezeigt. Dort erweist sich aber auch, daß AR(1)-Störgrößenprozesse im wesentlichen die einzigen AR-Prozesse sind, für die man dieses Ergebnis bei zunehmender Korrelation erster Ordnung der Störgrößen erwarten kann.

Trotz geringer relativer Effizienz wird FD gerade bei trendbehafteten Variablen oft als Alternative zu OLS empfohlen (GRANGER und NEWBOLD, 1974; NEWBOLD und DAVIES, 1978). Diese Empfehlung beruht aber nicht auf Effizienzüberlegungen in einem korrekt spezifizierten Modell, sondern auf der größeren Robustheit von FD gegen die fälschliche Konstatierung eines nicht vorhandenen Zusammenhangs zwischen abhängigen und unabhängigen Variablen (spurious regression, nonsense regression). Bei Trend in abhängigen und unabhängigen Variablen wird OLS (nicht aber FD) in der Regel auch ohne jeden Zusammenhang hohe Parameterwerte und ein hohes R^2 liefern (GRANGER und NEWBOLD, 1974). Diese geringere Anfälligkeit von FD gegen ungerechtfertigte "Schätzung" nichtvorhandener Zusammenhänge muß aber von Effizienzüberlegungen in einem korrekt spezifizierten Modell auseinandergelassen werden.

Eng verwandt mit Effizienzüberlegungen zu FD ist die in der Literatur viel diskutierte Frage, ob bei AR(1)-Störgrößen die erste

Beobachtung des für GLS implizit transformierten Modells berücksichtigt werden soll oder nicht (d.h. GLS versus COT). Obwohl in der Regel für $T \rightarrow \infty$ vernachlässigbar, bewirkt die erste transformierte Beobachtung bei festem T und trendbehafteten Daten einen großen Effizienzgewinn, wie in Abschnitt a) und von MAESHIRO (1976) gezeigt. PARK und MITCHELL (1980) bestätigen das mit Hilfe von Monte Carlo Experimenten auch für andere Trendvariable, sowohl für bekannte als auch für aus den Daten geschätzte ρ . Analoge Resultate liefert auch MAESHIRO (1980) für SUR (seemingly unrelated regresssion) Modelle. MAESHIRO (1979) gibt sogar für eine Reihe typischer Variablenverläufe konkrete Formeln für die Effizienzunterschiede an, und POIRIER (1978) stellt Methoden zur nachträglichen Berechnung von GLS-Schätzungen aus gegebenen COT-Schätzwerten bereit. TAYLOR (1981) schließlich diskutiert Bedingungen, unter denen die erste transformierte Beobachtung auch für $T \rightarrow \infty$ von Bedeutung bleibt.

Das in b) untersuchte Modell ähnelt der autoregressiven Beziehung

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t \quad (3.53)$$

für den Fall $\beta_2 = 1$. Das normierte zweite empirische Moment des Regressors verhält sich dann genauso wie im Modell (3.21). Wegen mangelnder Unabhängigkeit von Regressoren und Störgrößen ist die asymptotische Verteilungstheorie hier aber wesentlich komplizierter (ANDERSON, 1959; RAO, 1978; EVANS und SAVIN, 1981). Insbesondere ist (3.41) auch bei unabhängigen u_t nicht mehr asymptotisch normalverteilt (DICKEY und FULLER, 1981). WHITE (1958, 1959) und RAO (1961) schließlich untersuchen auch die asymptotische Verteilung von OLS in (3.53) für $|\beta_2| > 1$, d.h. den Fall des in der vorliegenden Arbeit ausgeschlossenen exponentiellen Trends.

Verallgemeinerungen der Ergebnisse aus b) bieten sich in verschiedene Richtungen an. Zunächst ist das Analogon von Satz (3.5) auch auf mehrere exogene Variable zu übertragen, von denen einige oder alle ARIMA-Prozesse sein mögen. Zum Beweis der asymptotischen Normalverteilung von GLS ist allein die Geltung von (2.22) für die Regressoren des transformierten Modells nachzuweisen.

Verallgemeinerungen von Satz (3.7) scheinen möglich für (i) ARIMA (p,d,q) -Prozesse mit $d > 1$, (ii) multiple statt bivariate Regressionen, und (iii) Störgrößenprozesse mit beliebiger stetiger Spektraldichte.

Die Beschränkung auf $d = 1$ bei den ARIMA (p,d,q) -Regressorprozessen ist allein durch die fundamentale beweistechnische Bedeutung von Lemma (3.6) bedingt. Die rege Diskussion des Grenzverhaltens empirischer Korrelationskoeffizienten bei ARIMA-Prozessen (WICHERN, 1973; ROY, 1977; ANDERSON, 1979; HASZA, 1980; ROY und LEFRANCOIS, 1981) läßt aber erwarten, daß (3.42) auch für ARIMA (p,d,q) -Prozesse mit $d > 1$ gezeigt werden kann. In diesem Fall wäre Satz (3.7) unmittelbar zu verallgemeinern.

Dagegen erscheint eine Verallgemeinerung der hier gebrauchten Beweistechnik auf multiple Regressionen nicht ohne weiteres möglich. Der Beweis von Satz (3.7) benutzt entscheidend die Darstellungen (3.26) bzw. (3.33) der OLS bzw. GLS-Schätzung für β_2 , eine nicht ohne weiteres auf multiple Regressionen übertragbare Möglichkeit.

Eine Erweiterung von Satz (3.7) auf Störgrößenprozesse mit beliebiger stetiger Spektraldichte schließlich bietet sich an auf dem Umweg über eine $AR(m)$ -Approximation, analog zu GRENANDER (1954). Jedoch ist die dortige Vorgangsweise wegen des stochastischen Charakters der hier vorliegenden Regressoren nicht unmittelbar zu übertragen, und andere Lösungsversuche haben noch nicht zum Ziel geführt.

4. TREND BEI FEHLERN IN DEN VARIABLEN

a) Modell und Schätzverfahren

Ökonometrische Modelle mit Fehlern in den Variablen stoßen in jüngster Zeit, nach einer langen Periode relativer Vernachlässigung, wieder auf zunehmendes Interesse. Ohnehin wurde die Bedeutung von Meßfehlern in nahezu allen ökonomischen Variablen kaum jemals bestritten, und standen vor allem technisch-statistische Probleme einer größeren Verbreitung im Wege.

In diesem Kapitel werden die Auswirkungen von Trend in den exogenen Variablen auf die asymptotischen Eigenschaften einiger Schätzverfahren in der funktionalen Variante des Fehler-in-den-Variablen Modells untersucht. Ausgangspunkt ist die Gleichung

$$y_t = \beta_1 z_{t1} + \dots + \beta_K z_{tK} + \varepsilon_t \quad (t=1,2,\dots) \quad , \quad (4.1)$$

oder kompakt in Matrixschreibweise (für einen Stichprobenumfang T)

$$y = Z \beta + \varepsilon \quad . \quad (4.2)$$

Dabei bezeichnet y in Fortführung der bisherigen Schreibweise den $T \times 1$ -Beobachtungsvektor der abhängigen Variablen, Z eine nichtstochastische $T \times K$ -Matrix ($T \geq K$; $\text{Rang}(Z)=K$) von Werten der exogenen Variablen, β den zu schätzenden $K \times 1$ -Parametervektor, und ε den $T \times 1$ -Störgrößenvektor der Regressionsgleichung.

Im Gegensatz zum Standardregressionsmodell seien gewisse Spalten

von Z nunmehr durch zufällige Beobachtungsfehler gestört, d.h. an Stelle von $Z = (Z^{(1)}:Z^{(2)})$ wird $X = (X^{(1)}:X^{(2)})$ beobachtet, mit $X^{(1)} = Z^{(1)}$, $X^{(2)} = Z^{(2)} + V^{(2)}$ und einer $T \times K_2$ -Zufallsmatrix $V^{(2)}$ ($0 \leq K_2 \leq K$). Mögliche Meßfehler bei y_t seien dagegen in ε_t enthalten. $\beta = (\beta^{(1)'}:\beta^{(2)'})'$ sei die analoge Zerlegung von β . Mit $V^{(1)} = 0$, $v = (V^{(1)}:V^{(2)})$ und $X = Z + V$ gilt dann

$$\begin{aligned} y &= (X^{(1)}:X^{(2)} - V^{(2)}) \begin{bmatrix} \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} + \varepsilon \\ &= (X^{(1)}:X^{(2)}) \begin{bmatrix} \beta^{(1)} \\ \beta^{(2)} \end{bmatrix} + \varepsilon - V^{(2)}\beta^{(2)} \\ &= X\beta + u \end{aligned} \tag{4.3}$$

mit

$$u = \varepsilon - V^{(2)}\beta^{(2)} \tag{4.4}$$

(4.3) entspricht der Standardnotation aus Kapitel 2 und liegt der OLS-Schätzung von β zugrunde.

Für die Zeilen $v_t^{(2)}$ von $V^{(2)}$ wird $E(v_t^{(2)}) = 0$ und

$$E\left\{ \begin{bmatrix} \varepsilon_t \\ v_t^{(2)} \end{bmatrix}, (\varepsilon_t, v_t^{(2)}) \right\} = \begin{bmatrix} \sigma_\varepsilon^2 & \omega' \\ \omega & \Omega \end{bmatrix} = \Sigma \text{ (regulär)} \tag{4.5}$$

vorausgesetzt. Zusätzlich seien $(\varepsilon_t, v_t^{(2)})$ und $(\varepsilon_s, v_s^{(2)})$ für $s \neq t$ unabhängig, mit identischer Verteilung. Das entspricht den Standard-

annahmen im vorliegenden Modellzusammenhang. Unter der zusätzlichen Standardannahme

$$\lim_{T \rightarrow \infty} T^{-1} Z'Z = Q^* \text{ (regulär)} \quad (4.6)$$

ist $\hat{\beta}^{\text{OLS}} = (X'X)^{-1} X'y$ wegen

$$E(T^{-1} X'u) = \begin{pmatrix} 0 \\ \omega - \Omega\beta^{(2)} \end{pmatrix} \neq 0 \quad (4.7)$$

(außer für eine Parametermenge von Lebesgue-Maß 0) damit im allgemeinen nicht mehr konsistent.

Zur Illustration betrachte man das weiter unten auch zur Simulation verwendete bivariate Modell

$$\begin{aligned} y_t &= \beta_1 + \beta_2 z_{t2} + \varepsilon_t \\ &= \beta_1 + \beta_2 x_{t2} + u_t \quad (t=1,2,\dots) \end{aligned} \quad (4.8)$$

mit $u_t = \varepsilon_t - \beta_2 v_t$. Hier ist $Z^{(1)} = X^{(1)} = (1,1,\dots,1)'$, $Z^{(2)} = (z_{12}, z_{22}, \dots, z_{T2})'$, $X^{(2)} = Z^{(2)} + (v_1, \dots, v_T)'$, $\beta^{(1)} = \beta_1$, $\beta^{(2)} = \beta_2$, $K = 2$ und $K_2 = 1$. Zur Vereinfachung sei weiter $E(v_t \varepsilon_t) = 0$.

In diesem einfachen Modell reduziert sich (4.6) auf $0 < m_{xx} := \lim_{T \rightarrow \infty} T^{-1} \Sigma (z_{t2} - \bar{z}_2)^2 < \infty$. Daraus folgt aber

$$0 < m_{xx} := \text{plim}_{T \rightarrow \infty} T^{-1} \Sigma (x_{t2} - \bar{x}_2)^2 = m_{zz} + \sigma_v^2 \quad (4.9)$$

und

$$\begin{aligned}
\text{plim}_{T \rightarrow \infty} (\hat{\beta}_2^{\text{OLS}} - \beta_2) &= \text{plim}_{T \rightarrow \infty} \frac{T^{-1} \sum (x_{t2} - \bar{x}_2) u_t}{T^{-1} \sum (x_{t2} - \bar{x}_2)^2} \\
&= - \frac{\beta_2 \sigma_v^2}{m_{zz} + \sigma_v^2} \\
&\neq 0
\end{aligned} \tag{4.10}$$

(außer für $\beta_2 = 0$), d.h. die Inkonsistenz von OLS. Dieses Ergebnis ist wohlbekannt (für neuere Diskussionen siehe etwa SCHÖNFELD, 1971, S. 105; THEIL, 1971, S. 608; SCHNEEWEISS, 1971, Abschnitt 7.2.1; 1981, S.48; SCHNEEWEISS und MITTAG, 1984, Abschnitt 3.2.1) und zeigt insbesondere, daß OLS den Absolutbetrag des Steigungskoeffizienten in diesem einfachen Modell systematisch unterschätzt. Diese asymptotische Verzerrung nimmt aber mit wachsendem Grenzwert des zweiten empirischen Moments der unabhängigen Variablen ab, was als Plausibilitätsüberlegung zu der weiter unten zu zeigenden Konsistenz von OLS im Trendfall (d.h. unendlichen Grenzwerten der zweiten empirischen Momente) herangezogen werden kann.

Alternative und auch unter (4.6) konsistente Schätzverfahren unterscheiden sich vor allem durch die zur Schätzung von β herangezogene Zusatzinformation. Dabei soll für die vorliegende Untersuchung offen bleiben, aus welchen Quellen diese Kenntnisse stammen. Zwei dieser Verfahren, die sogenannte "Korregierte Kleinst-Quadrate Methode" (im weiteren CLS für "Corrected Least Squares"; siehe MADANSKY, 1959; THEIL, 1971, S. 614; SCHNEEWEISS, 1976; KETELLAPPER, 1982, 1983) und GLS (siehe SCHÖNFELD, 1971, Kap. 7) werden im weiteren mit OLS verglichen.

CLS setzt die Kenntnis bzw. konsistente Schätzbarkeit von Ω (der in (4.5) definierten Kovarianzmatrix der Meßfehler der exogenen Variablen) voraus. Mit

$$\tilde{\Omega} = E(v_t^{(1)}, v_t^{(2)})', (v_t^{(1)}, v_t^{(2)}) = \begin{bmatrix} 0 & 0 \\ 0 & \Omega \end{bmatrix}$$

ist diese Schätzung gegeben durch

$$\hat{\beta}^{\text{CLS}} = (X'X - T\tilde{\Omega})^{-1} X'y \quad (4.11)$$

Offenbar ist $\hat{\beta}^{\text{CLS}}$ bei fehlerfreien Beobachtungen mit OLS identisch und für $\omega = 0$ auch konsistent (siehe SCHNEEWEISS, 1976).

In SCHNEEWEISS und MITTAG (1984) wird CLS auch als V-Schätzung, GLS als P-Schätzung bezeichnet.

Die GLS-Schätzung $\hat{\beta}^{\text{GLS}}$ für β beruht auf der Kenntnis von Σ (bis auf eine skalare multiplikative Konstante) und ist definiert durch

$$\sum_{t=1}^T (\varepsilon_t, v_t^{(2)}) \Sigma^{-1} (\varepsilon_t, v_t^{(2)})' = \min! \quad (4.12)$$

mit der Nebenbedingung

$$y = X^{(1)} \hat{\beta}^{(1)\text{GLS}} + (X^{(2)} - V^{(2)}) \hat{\beta}^{(2)\text{GLS}} + \varepsilon \quad (4.13)$$

Dieses Minimierungsproblem, mit den Variablen ε , $V^{(2)}$ und $\hat{\beta}^{\text{GLS}} = (\hat{\beta}^{(1)\text{GLS}}, \hat{\beta}^{(2)\text{GLS}})'$, bleibt unter skalarer Multiplikation von Σ invariant. Im allgemeinen führt es nicht zu einem geschlossenen Ausdruck

für $\hat{\beta}^{\text{GLS}}$. Bei normalverteilten Fehlern und Störgrößen liefert $\hat{\beta}^{\text{GLS}}$ zugleich die Maximum-Likelihood Schätzung für β . Einzelheiten dazu sind etwa bei SCHÖNFELD (1971, S. 114-118) nachzulesen. Wie man weiter sofort sieht, wird auch durch das in (4.12) und (4.13) gegebene Minimierungsproblem für $V^{(2)} = \text{leer}$ (keine Fehler in den exogenen Variablen) gerade die OLS-Schätzung $\hat{\beta}^{\text{OLS}}$ definiert.

Konsistenz und asymptotische Normalverteilung unter (4.6) werden von SCHNEEWEISS (1976) für $\hat{\beta}^{\text{CLS}}$ und von MALINVAUD (1981, Kapitel 10) für $\hat{\beta}^{\text{GLS}}$ gezeigt. Im weiteren stehen analoge Überlegungen für den Fall trendbehafteter exogener Variablen im Mittelpunkt. Dabei wird an Stelle von (4.6) lediglich das Analogon von (2.7) und (2.8) vorausgesetzt, in der Notation dieses Kapitels

$$\max_{1 \leq t \leq T} z_{tk}^2 / \sum_{t=1}^T z_{tk}^2 \longrightarrow 0 \quad (k=1, \dots, K) \quad (4.14)$$

und

$$D_T^{-1} Z'Z D_T^{-1} \longrightarrow Q \text{ (regulär)} \quad (4.15)$$

mit $D_T = \text{diag}(\|z_{\cdot 1}\|, \dots, \|z_{\cdot K}\|)$. Wegen Satz (2.1) folgt daraus

$$\max_{1 \leq t \leq T} \zeta_{tt} \longrightarrow 0 \quad (4.16)$$

mit ζ_{tt} als t 'tem Diagonalelement in $Z(Z'Z)^{-1}Z'$. Entsprechend der Vereinbarung (2.20) ist ferner Trend in der i 'ten exogenen Variablen definiert durch

$$T^{-1/2} \|z_{\cdot i}\| \rightarrow \infty \quad (4.17)$$

Die Annahmen (4.14) und (4.15) bilden die Grundlage der asymptotischen Resultate dieses Kapitels und auch von Kapitel 5 zu Simultanen Gleichungssystemen. Obwohl per se nur Verallgemeinerung von (4.6), sind damit im Verein mit (4.17) einige unerwartete Implikationen verknüpft. Bei mehr als einer trendbehafteten Variablen unter den Regressoren beinhaltet (4.15) Restriktionen zu den Abweichungen vom Trend, die bestimmte Variablenverläufe ausschließen. Zur Illustration betrachte man den Fall $K = 2$ und

$$z_{t1} = t ; \quad z_{t2} = t + \cos(\pi t) \quad (t=1,2,\dots) \quad (4.18)$$

Beide Variablen erfüllen (4.14) und (4.17), aber

$$D_T^{-1} Z'Z D_T^{-1} \rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ,$$

eine singuläre Matrix. Damit ist (4.15) verletzt und sind Fälle wie (4.18) durch die folgenden Sätze nicht erfaßt.

(4.15) fordert damit etwa, daß bei linearem Trend und mehr als einer trendbehafteten Variablen unter den exogenen Variablen des Modells die Abweichungen vom Trend mit höchstens einer Ausnahme selbst trendbehaftet sein müssen (d.h. nicht beschränkt bleiben dürfen). Bei unterjährigen Daten sind derartige Variablenverläufe

mit trendabhängigen additiven Abweichungen vom Trend bei Saisonbereinigungsverfahren mit multiplikativer Saisonfigur immer implizit unterstellt, aber auch in Modellen mit additiver Saisonfigur zuweilen explizit zugelassen (WINCKLER, 1970; KRÄMER, 1979). Ein Beispiel ist

$$z_{1t} = t; z_{2t} = t + t \cos(\pi t) \quad . \quad (4.19)$$

Hier gilt

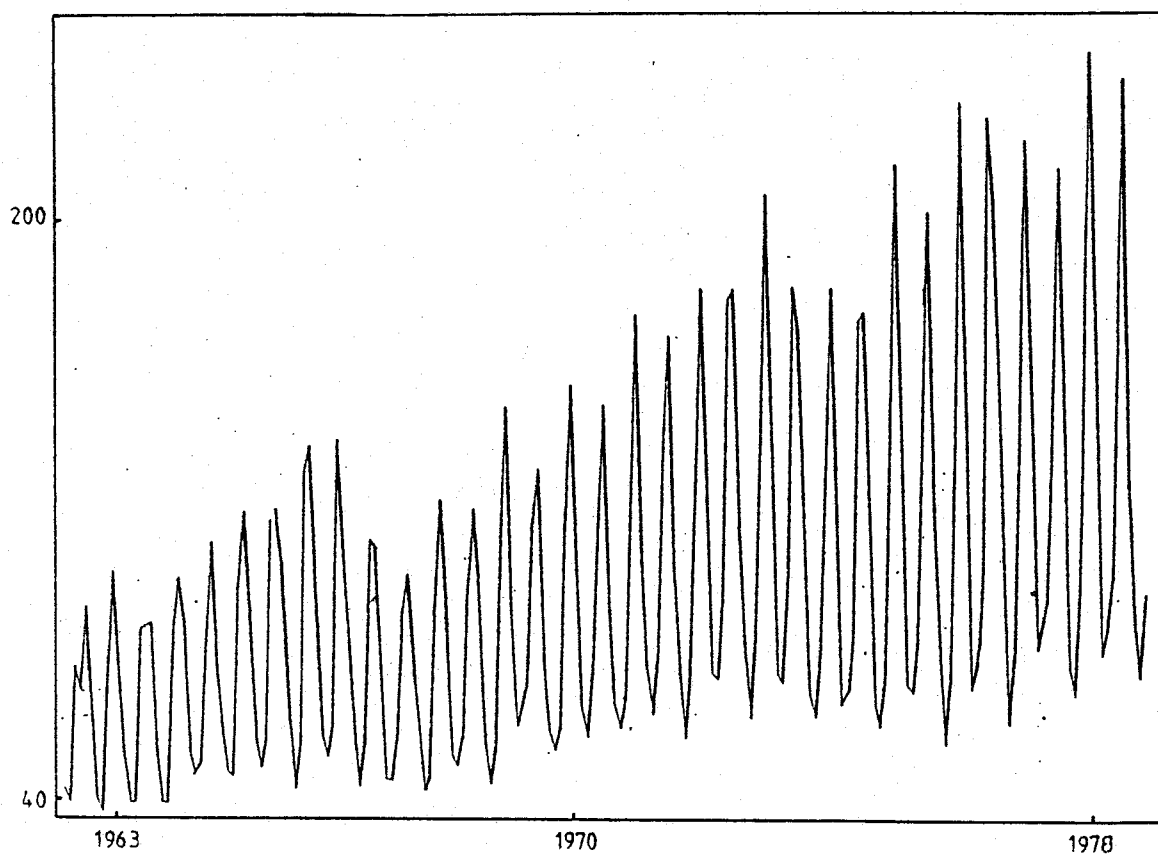
$$D_T^{-1} Z' T D_T^{-1} \longrightarrow \begin{bmatrix} 1 & 2^{-1/2} \\ 2^{-1/2} & 1 \end{bmatrix}$$

und (4.15) ist erfüllt. Auch bei polynomialen Trends abweichender Ordnung sind (4.14), (4.15) und (4.17) durchaus vereinbar (GRANDNER und ROSENBLATT, 1957, S. 245; EICKER, 1967, S. 64).

Letztendlich muß jedoch die ökonomische Realität über die Brauchbarkeit der hier getroffenen Annahmen entscheiden. Daher werden im Anhang dieser Arbeit eine Reihe trendbehafteter Zeitreihen auf ihre Verträglichkeit mit obigen Annahmen untersucht. Als vorläufige Illustration gibt Schaubild (4.1a), entnommen aus JÖCKEL (1982, S. 371), einen typischen hier zugelassenen Kurvenverlauf wieder, entstanden aus von HEILER (1980) untersuchten Monatsdaten zum Nettoproduktionsindex der westdeutschen Bekleidungsindustrie.

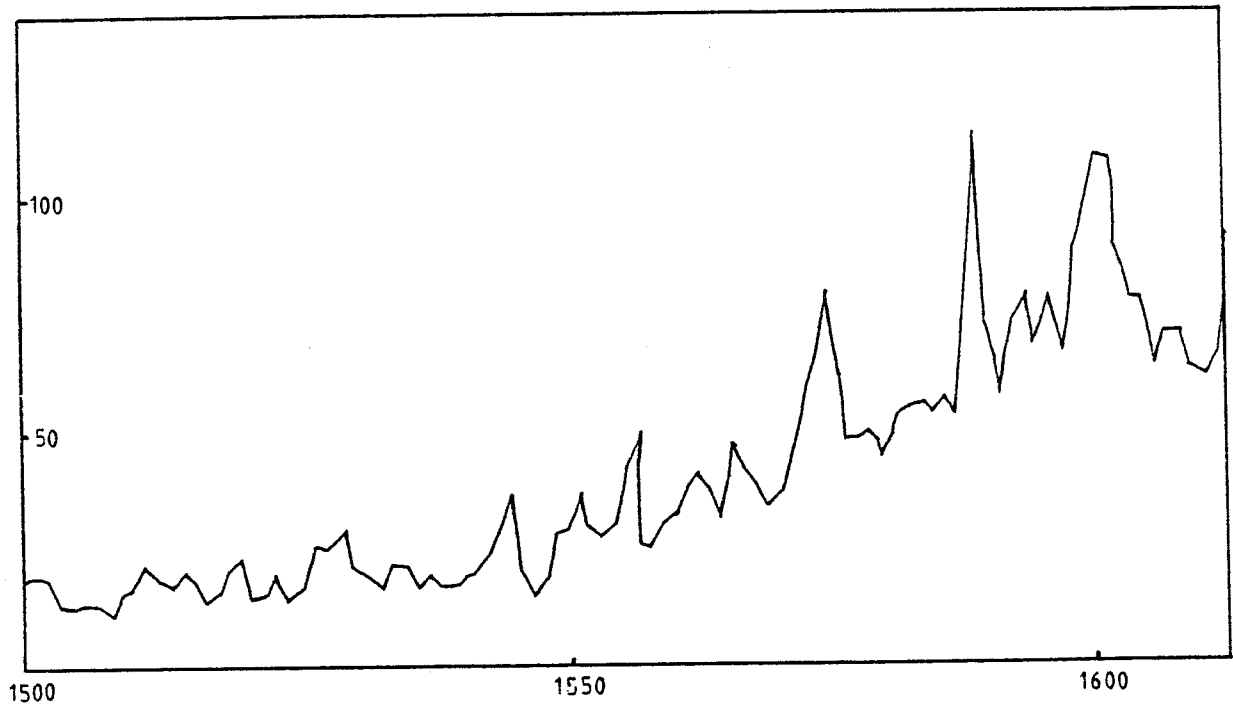
Bei Jahresdaten kann man derartige streng periodische Trendabweichungen natürlich nicht erwarten, aber auch hier sind zunehmende Abweichungen von einem langfristigen Trend nicht ausgeschlossen. Schaubild (4.1b) gibt zur Illustration die ersten 120 Werte einer

SCHAUBILD (4.1a): Monatsdaten zum Nettoproduktionsindex der westdeutschen Bekleidungsindustrie



der längsten bekannten ökonomischen Zeitreihen wieder, des von Beveridge zusammengetragenen Weizenpreis-Index' (entnommen aus ANDERSON, 1971, S. 623-626). Auch hier sind Zyklen mit zunehmender Amplitude unschwer auszumachen.

SCHAUBILD (4.1b): Der Beveridge Weizenpreis-Index



Vorerst sei jedoch die Diskussion der Realitätsnähe unserer Annahmen zurückgestellt und nochmals auf deren implizite Konsequenzen hingewiesen.

b) Konsistenz und Effizienz von OLS

Im weiteren seien trendfreie Variable durch ein Superskript a, trendbehaftete Variable durch ein Superskript b gekennzeichnet. Z sei nach den Kriterien exakt/fehlerbehaftet und trendfrei/Trend zerlegt in

$$Z = (Z^{(1a)} : Z^{(1b)} : Z^{(2a)} : Z^{(2b)}) . \quad (4.20)$$

Damit faßt etwa $Z^{(1a)}$ die trendfreien und exakt gemessenen Spalten von Z zusammen. Entsprechend seien auch V , X , D_T und Σ zerlegt. So wird etwa (4.5) in ausführlicher Schreibweise zu

$$\begin{aligned} \Sigma &= E \left\{ \begin{bmatrix} \varepsilon_t \\ v_t^{(2a)} \\ v_t^{(2b)} \end{bmatrix} \begin{bmatrix} \varepsilon_t & v_t^{(2a)} & v_t^{(2b)} \end{bmatrix} \right\} \\ &= \begin{bmatrix} \sigma_\varepsilon^2 & \omega^{(a)} & \omega^{(b)} \\ \omega^{(a)} & \Omega^{(aa)} & \Omega^{(ab)} \\ \omega^{(b)} & \Omega^{(ba)} & \Omega^{(bb)} \end{bmatrix} \end{aligned} \quad (4.21)$$

SATZ (4.1): $Z^{(2a)}$ leer (alle fehlerhaften Variablen enthalten Trend) $\Rightarrow \hat{\beta}^{OLS}$ ist (schwach) konsistent. Unter der Zusatzvoraussetzung

$$T^{-1/2} D_T^{(a)} \longrightarrow \bar{D}^{(a)} \text{ (regulär)} \quad (4.22)$$

(eine Verschärfung von (4.15)), und abgesehen von einer Parametermenge vom Maß 0, gilt auch die Umkehrung.

BEWEIS: Sei zunächst $Z^{(2a)}$ leer, d.h.

$$D_T = \begin{vmatrix} D_T^{(1a)} & & & \\ & D_T^{(1b)} & 0 & \\ & 0 & & \\ & & & D_T^{(2b)} \end{vmatrix} .$$

Dann gilt

$$\begin{aligned} \hat{\beta}^{\text{OLS}} - \beta &= (X'X)^{-1} X'u \\ &= D_T^{-1} D_T (X'X)^{-1} D_T D_T^{-1} X'u \end{aligned} \quad (4.23)$$

mit

$$D_T^{-1} X'u = D_T^{-1} Z'u + D_T^{-1} V'u = O_p(1) . \quad (4.24)$$

Die letzte Behauptung folgt dabei unmittelbar aus $D_T^{-1} V'u \xrightarrow{P} 0$ und der Konstanz der Varianz aller Komponenten von $D_T^{-1} Z'u$. Wegen $Z^{(2a)} = \text{leer}$ prüft man weiterhin leicht nach, daß

$$\text{plim}_{T \rightarrow \infty} D_T^{-1} X'X D_T^{-1} = \lim_{T \rightarrow \infty} D_T^{-1} Z'Z D_T^{-1} = Q , \quad (4.25)$$

woraus wegen $D_T (X'X)^{-1} D_T \xrightarrow{P} Q^{-1} < \infty$, $D_T^{-1} \rightarrow 0$ und (4.23) die Konsistenz von OLS sofort folgt.

Sei nun umgekehrt $Z^{(2a)}$ nicht leer, und zusätzlich existiere $\bar{D}^{(a)} = \lim_{T \rightarrow \infty} T^{-1/2} D_T^{(a)}$ (regulär). Dann gilt (4.25) nicht mehr, aber

$$\text{plim}_{T \rightarrow \infty} D_T^{-1} X'X D_T^{-1} = Q + R \quad (4.26)$$

ist weiterhin regulär. Dabei ist R eine Blockdiagonalmatrix mit

$\bar{D}^{(2a)-1} \Omega^{(aa)} \bar{D}^{(2a)-1}$ als einzigem von 0 verschiedenen Block. Man arrangiere ferner die Spalten von Z und X dergestalt um, daß die trendfreien Komponenten zusammenhängende Teilmatrizen bilden, d.h. $Z = (Z^{(a)} : Z^{(b)})$ und $X = (X^{(a)} : X^{(b)})$. Mit diesen Vereinbarungen läßt sich (4.23) schreiben als

$$\hat{\beta}^{OLS} - \beta = \{T^{1/2} D_T^{-1}\} \{D_T (X'X)^{-1} D_T\} \{T^{-1/2} D_T^{-1} X'u\}, \quad (4.27)$$

wobei

$$T^{1/2} D_T^{-1} \longrightarrow \begin{bmatrix} \bar{D}^{(a)-1} & 0 \\ 0 & 0 \end{bmatrix}, \quad (4.28)$$

$$D_T (X'X)^{-1} D_T \xrightarrow{P} (Q+R)^{-1} < \infty, \quad (4.29)$$

und

$$\begin{aligned} T^{-1/2} D_T^{-1} X'u &= T^{-1/2} D_T^{-1} Z'u + T^{-1/2} D_T^{-1} V'u \\ &\xrightarrow{P} \lim(T^{1/2} D_T^{-1}) \text{plim}(T^{-1} V'u) \\ &= \begin{bmatrix} \bar{D}^{(a)-1} & \text{plim}(T^{-1} V^{(a)'u}) \\ 0 & \end{bmatrix} \\ &\neq 0. \end{aligned} \quad (4.30)$$

Die letzte Ungleichung, aus der die Inkonsistenz von OLS unmittelbar folgt, ergibt sich dabei aus

$$\text{plim}_{T \rightarrow \infty} V^{(2a)'u} = \omega^{(a)} - \Omega^{(aa)} \beta^{(2a)} - \Omega^{(ab)} \beta^{(2b)} \neq 0 \quad (4.31)$$

(bis auf eine Parametermenge vom Lebesgue-Maß 0). q.e.d.

(4.22) bedeutet in der Tat eine Verschärfung von (4.15), wie etwa das Beispiel $z_{it} = t^{-1/2}$ zeigt. Diese Folge erfüllt (4.14) und steht auch (4.15) nicht im Weg, aber $\sum_{t=1}^T z_{it}^2 = O(\ln(T))$, d.h. $T^{-1/2} \|z_{\cdot i}\| \rightarrow 0$ und $\bar{D}^{(a)}$ ist singulär. Allerdings ist die Regularitätsbedingung für $\bar{D}^{(a)}$ allein durch die oben gewählte Beweistechnik bedingt und möglicherweise überflüssig.

Wie der Beweis ferner zeigt, wird (4.14) für die Konsistenz von OLS an keiner Stelle gebraucht. Diese ist damit durch Satz (4.1) auch für nicht langsam wachsende Regressoren, etwa exponentielle Trends, gezeigt, sofern nur die restlichen Bedingungen erfüllt sind. Zur Herleitung der asymptotischen Verteilung von OLS spielt (4.14) aber eine zentrale Rolle. Diese asymptotische Verteilung wird im weiteren nur für konsistente OLS-Schätzungen betrachtet, da andernfalls eine OLS-Schätzung wenig sinnvoll erscheint.

SATZ (4.2): Sei $Z^{(2a)}$ leer, und es gelte (4.14) und (4.15). Dann gilt

$$D_T(\hat{\beta}^{\text{OLS}} - \beta) \xrightarrow{D} N(0, \sigma_u^2 Q^{-1}), \quad (4.32)$$

und diese Grenzverteilung ist identisch mit der von $D_T(\hat{\beta}^{\text{GLS}} - \beta)$ und $D_T(\hat{\beta}^{\text{CLS}} - \beta)$.

BEWEIS: Man hat

$$\begin{aligned}
D_T(\hat{\beta}^{\text{OLS}} - \beta) &= D_T(X'X)^{-1}X'u \\
&= D_T(Z'Z)^{-1}Z'u + \{D_T(X'X)^{-1}D_T - D_T(Z'Z)^{-1}D_T\} D_T^{-1}X'u \\
&\quad + D_T(Z'Z)^{-1}D_T\{D_T^{-1}X'u - D_T^{-1}Z'u\} . \quad (4.33)
\end{aligned}$$

Die Ausdrücke in geschweiften Klammern verschwinden dabei für $T \rightarrow \infty$ wegen (4.25) und

$$D_T^{-1}X'u - D_T^{-1}Z'u = D_T^{-1}V'u \xrightarrow{P} 0 .$$

Wegen $D_T^{-1}X'u = o_p(1)$ und $D_T(Z'Z)^{-1}D_T = o_p(1)$ verschwinden damit alle Summanden in (4.33) außer $D_T(Z'Z)^{-1}Z'u$. Dieser Ausdruck ist aber der normalisierte Schätzfehler im Modell $y = Z\beta + u$, für den wegen Satz (2.3) und (2.28) gilt, daß

$$D_T(Z'Z)^{-1}Z'u \xrightarrow{D} N(0, \sigma_u^2 Q^{-1}) . \quad (4.34)$$

Damit ist (4.32) bewiesen. Zum Beweis des zweiten Teils von Satz (4.2) wird gezeigt, daß

$$D_T(\hat{\beta}^{\text{OLS}} - \hat{\beta}^{\text{GLS}}) \xrightarrow{P} 0 \quad (4.35)$$

und

$$D_T(\hat{\beta}^{\text{OLS}} - \hat{\beta}^{\text{CLS}}) \xrightarrow{P} 0 . \quad (4.36)$$

Dabei folgt (4.36) sofort aus

$$\begin{aligned}
D_T(\hat{\beta}^{\text{OLS}} - \hat{\beta}^{\text{CLS}}) &= D_T \{(X'X)^{-1} - (X'X - T\tilde{\Omega})^{-1}\} X'y \\
&= \{D_T(X'X)^{-1}D_T - (D_T^{-1}(X'X - T\tilde{\Omega})D_T^{-1})^{-1}\} D_T^{-1}X'y ,
\end{aligned}$$

$$D_T^{-1}X'y = o_p(1) \text{ und}$$

$$\{D_T(X'X)^{-1}D_T - (D_T(X'X - T\tilde{\Omega})D_T^{-1})^{-1} \xrightarrow{P} 0 \quad (4.37)$$

Zum Beweis von (4.37) beachte man, daß

$$D_T^{-1}(X'X - T\tilde{\Omega})D_T^{-1} = D_T^{-1}X'X D_T^{-1} - T^{1/2}D_T^{-1}\tilde{\Omega}T^{1/2}D_T^{-1} \quad ,$$

wobei $D_T^{-1}X'X D_T^{-1} \xrightarrow{P} Q$ (wegen (4.25)), und

$$T^{1/2}D_T^{-1}\tilde{\Omega}T^{1/2}D_T^{-1} \xrightarrow{P} 0$$

(da für diejenigen Diagonalelemente von D_T , die zu den von 0 verschiedenen Zeilen und Spalten von $\tilde{\Omega}$ gehören, nach Voraussetzung $D_{T,ii}/T^{1/2} \rightarrow \infty$, d.h. $T^{1/2}/D_{T,ii} \rightarrow 0$).

Der Beweis von (4.35) erfordert einige zusätzliche Überlegungen. Dazu sei $(\beta^{(1)}, \beta^{(2)})'$ die in (4.3) vereinbarte Zerlegung von β entsprechend der Aufteilung $X = (X^{(1)}; X^{(2)})$ in korrekt gemessene und gestörte exogene Variable. Nach SCHÖNFELD (1971, S. 117) ist dann $\hat{\beta}^{(2)GLS}$ gegeben durch

$$(W - \mu\Sigma) \begin{bmatrix} -1 \\ \hat{\beta}^{(2)GLS} \end{bmatrix} = 0 \quad (4.38)$$

mit

$$W = \begin{bmatrix} \tilde{y}'\tilde{y} & \tilde{y}'\tilde{X}^{(2)} \\ \tilde{X}^{(2)'}\tilde{y} & \tilde{X}^{(2)'}\tilde{X}^{(2)} \end{bmatrix} \quad ,$$

$$\tilde{y} = y - X^{(1)}(X^{(1)'}, X^{(1)})^{-1}X^{(1)'}y \quad ,$$

$$\tilde{X}^{(2)} = X^{(2)} - X^{(1)}(X^{(1)'}, X^{(1)})^{-1}X^{(1)'}X^{(2)} \quad ,$$

μ kleinster Eigenwert von $P W P'$,

$$P \text{ regulär mit } P'P = \Sigma^{-1} \text{ und } \Sigma = \begin{vmatrix} \sigma_\varepsilon^2 & -\omega' \\ -\omega & \Omega \end{vmatrix} .$$

Daraus folgt sofort

$$\hat{\beta}^{(2)\text{GLS}} = (\tilde{X}^{(2)'} \tilde{X}^{(2)} - \mu\Omega)^{-1} (\tilde{X}^{(2)'} y - \mu\omega) . \quad (4.39)$$

Andererseits gilt

$$\hat{\beta}^{(2)\text{OLS}} = (\tilde{X}^{(2)'} \tilde{X}^{(2)})^{-1} \tilde{X}^{(2)'} \tilde{y} , \quad (4.40)$$

d.h.

$$\begin{aligned} & D_T^{(2)} (\hat{\beta}^{(2)\text{GLS}} - \hat{\beta}^{(2)\text{OLS}}) \\ &= \{D_T^{(2)-1} (\tilde{X}^{(2)'} \tilde{X}^{(2)} - \mu\Omega) D_T^{(2)-1}\}^{-1} D_T^{(2)-1} (\tilde{X}^{(2)'} \tilde{y} - \mu\omega) \\ &\quad - \{D_T^{(2)-1} \tilde{X}^{(2)'} \tilde{X}^{(2)} D_T^{(2)-1}\}^{-1} D_T^{(2)-1} \tilde{X}^{(2)'} y \\ &= \{D_T^{(2)-1} (\tilde{X}^{(2)'} \tilde{X}^{(2)} - \mu\Omega) D_T^{(2)-1}\}^{-1} - \{D_T^{(2)-1} \tilde{X}^{(2)'} \tilde{X}^{(2)} D_T^{(2)-1}\}^{-1} \\ &\quad \cdot D_T^{(2)-1} (\tilde{X}^{(2)'} \tilde{y} - \mu\omega) \\ &\quad + \{D_T^{(2)-1} \tilde{X}^{(2)'} \tilde{X}^{(2)} D_T^{(2)-1}\}^{-1} \{D_T^{(2)-1} \mu\omega\} \\ &\xrightarrow{P} 0 . \end{aligned} \quad (4.41)$$

Zum Beweis dieser letzten Behauptung beachte man, daß

$$\mu = \min_a \frac{a' P W P' a}{a' a} = \min_b \frac{b' W B}{b' P^{-1} (P^{-1})' b}$$

mit von Null verschiedenen Vektoren a und $b = P'a$. Für $b = (1, -\beta^{(2)'})'$ liefert dies

$$\mu \leq \|u - X^{(1)} (X^{(1)'} X^{(1)})^{-1} X^{(1)'} u\|^2 / b' \Sigma b = o(T) = o(D_T^{(2)})$$

(komponentenweise), d.h. in (4.41) verschwinden die Ausdrücke in geschweiften Klammern für $T \rightarrow \infty$. Wegen stochastischer Beschränktheit der übrigen Ausdrücke folgt daraus aber die Behauptung.

Es bleibt noch zu zeigen, daß auch

$$D_T^{(1)} (\hat{\beta}^{(1)OLS} - \hat{\beta}^{(1)GLS}) \xrightarrow{P} 0. \quad (4.42)$$

Dazu beachte man (siehe SCHÖNFELD, 1971, S. 116), daß

$$\hat{\beta}^{(1)GLS} = (X^{(1)}, X^{(1)})^{-1} X^{(1)}, (y: X^{(2)}) \begin{bmatrix} 1 \\ -\hat{\beta}^{(2)GLS} \end{bmatrix},$$

während

$$\hat{\beta}^{(1)OLS} = (X^{(1)}, X^{(1)})^{-1} X^{(1)}, (y: X^{(2)}) \begin{bmatrix} 1 \\ -\hat{\beta}^{(2)OLS} \end{bmatrix},$$

d.h.

$$\begin{aligned} & D_T^{(1)} (\hat{\beta}^{(1)OLS} - \hat{\beta}^{(1)GLS}) \\ &= D_T^{(1)} (X^{(1)}, X^{(1)})^{-1} D_T^{(1)} D_T^{(1)-1} X^{(1)}, X^{(2)} D_T^{(2)-1} D_T^{(2)} \cdot (\hat{\beta}^{(2)OLS} - \hat{\beta}^{(2)GLS}). \end{aligned}$$

Da $Q^{-1} = \text{plim}_{T \rightarrow \infty} D_T (X'X)^{-1} D_T$ regulär, existieren auch

$$\text{plim}_{T \rightarrow \infty} D_T^{(1)} (X^{(1)}, X^{(1)})^{-1} D_T^{(1)} < \infty$$

und

$$\text{plim}_{T \rightarrow \infty} D_T^{(1)-1} X^{(1)}, X^{(2)} D_T^{(2)-1} < \infty,$$

woraus (4.42) wegen $D_T^{(2)} (\hat{\beta}^{(2)OLS} - \hat{\beta}^{(2)GLS}) \xrightarrow{P} 0$ sofort folgt und was den Beweis des Satzes komplettiert. q.e.d.

Zur Konstruktion von OLS-gestützten (asymptotischen) Signifikanztests und Konfidenzintervallen bleibt noch die Konsistenz der OLS-Schätzung für σ_u^2 zu zeigen. Wegen Korrelation von Regressoren und Störgrößen ist Satz (2.4) hier nicht anwendbar.

SATZ (4.3): Unter (4.14) und (4.15), und für $Z^{(2a)} = \text{leer}$, gilt

$$s^2 = (y - X\hat{\beta}^{\text{OLS}})'(y - X\hat{\beta}^{\text{OLS}})/(T-K) \xrightarrow{P} \sigma_u^2. \quad (4.43)$$

BEWEIS: Wegen

$$s^2 = \frac{1}{T-K} u'u - \frac{1}{T-K} u'X(X'X)^{-1}X'u$$

und $u'u/(T-K) \xrightarrow{P} \sigma_u^2$ reicht es wie im Beweis von Satz (2.4) zu zeigen, daß $u'X(X'X)^{-1}X'u/(T-K) \xrightarrow{P} 0$. Dies wiederum folgt wegen

$$(T-K)^{-1}u'X(X'X)^{-1}X'u = (T-K)^{-1/2}u'XD_T^{-1}\{D_T(X'X)^{-1}D_T\}D_T^{-1}X'u(T-K)^{-1/2} \quad (4.44)$$

und $D_T(X'X)^{-1}D_T = O_p(1)$ aus

$$(T-K)^{-1/2}u'XD_T^{-1} \xrightarrow{P} 0. \quad (4.45)$$

Zum Beweis von (4.45) zerlege man $X = (X^{(a)}:X^{(b)})$ in trendfreie und trendbehaftete Spalten, mit

$$\begin{aligned} (T-K)^{-1/2}u'XD_T^{-1} &= ((T-K)^{-1/2}u'X^{(a)}D_T^{(a)-1} : \\ &\quad (T-K)^{-1/2}u'Z^{(b)}D_T^{(b)-1}) \\ &\quad + (T-K)^{-1}u'V^{(b)}(D_T^{(b)})/(T-K)^{1/2})^{-1}. \end{aligned} \quad (4.46)$$

Nach Voraussetzung ist $X^{(a)} = Z^{(a)}$ fehlerfrei, d.h.
 $u'X^{(a)}D_T^{(a)-1} = O_p(1)$ und $(T-K)^{-1/2}u'X^{(a)}D_T^{(a)-1} \xrightarrow{p} 0$. Das gleiche
 Argument liefert auch die Konvergenz gegen Null des zweiten Terms.
 Der letzte Term in (4.46) schließlich verschwindet wegen $(T-K)^{-1}$
 $(T-K)^{-1}u'V^{(b)} = O_p(1)$ und $(D_T^{(b)})/(T-K)^{1/2} \xrightarrow{p} 0$. q.e.d.

Eine letzte Zusatzüberlegung wird durch den Umstand nötig, daß
 D_T als Funktion der unbeobachtbaren Komponenten von Z selbst ebenfalls
 nicht beobachtbar ist. Stattdessen kann jedoch \tilde{D}_T verwendet werden,
 mit

$$\tilde{D}_{T,ii} = \left(\sum_{t=1}^T x_{ti}^2 \right)^{1/2}. \quad (4.47)$$

Unter den Voraussetzungen von Satz (4.2) gilt

$$\tilde{D}_T D_T^{-1} \xrightarrow{p} I_K, \quad (4.48)$$

d.h. $\tilde{D}_T(\hat{\beta}^{OLS} - \beta) = \tilde{D}_T D_T^{-1} D_T(\hat{\beta}^{OLS} - \beta)$ und $D_T(\hat{\beta}^{OLS} - \beta)$ haben die
 gleiche Grenzverteilung. Als Fazit von Satz (4.2) und Satz (4.3) er-
 hält man damit, daß auch bei Fehlern in den Variablen, sofern alle
 fehlerbehafteten Variablen einen Trend aufweisen und die Annahmen dieses
 Kapitels gelten, verfahren werden kann wie im Standardregressionsmodell.

c) Simulation

Analog zu Kapitel 3 erfolgt nun eine Monte-Carlo Überprüfung der Aussagekraft obiger Resultate in endlichen Stichproben. Alle Experimente wurden wiederum mit dem Sperry-Univac 1100 Rechner des Instituts für Höhere Studien durchgeführt, unter Verwendung eigener Fortran-Programme und mit den gleichen Bibliotheks-Routinen zur Erzeugung von Pseudo-Zufallszahlen wie in Kapitel 3.

Grundlage aller Experimente ist das bivariate Modell (4.8), mit $y_t = \beta_1 + \beta_2 z_{t2} + \varepsilon_t = \beta_1 + \beta_2 x_{t2} + u_t$, $x_{t2} = z_{t2} + v_t$, $u_t = \varepsilon_t - \beta_2 v_t$, und

$$E(v_t \varepsilon_t) = 0 . \quad (4.49)$$

Dieses einfache Modell erlaubt eine im Vergleich zu multiplen Regressionen dichtere Überdeckung des relevanten Parameterraumes und insbesondere auch eine Konfrontation von Monte Carlo Ergebnissen mit analytischen endliche-Stichproben Resultaten der einschlägigen Literatur. Wie in 3 c) stehen dabei wiederum die Schätzungen für den Steigungskoeffizienten β_2 im Mittelpunkt.

Ein weiterer Vorteil dieses einfachen Modells ist die Existenz geschlossener Ausdrücke für alle betrachteten Schätzverfahren. OLS liefert wie in (3.26)

$$\hat{\beta}_2^{\text{OLS}} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2} , \quad (4.50)$$

die Formel (4.11) zur Berechnung von $\hat{\beta}^{\text{CLS}}$ reduziert sich auf

$$\hat{\beta}_2^{\text{CLS}} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2 - T\sigma_v^2}, \quad (4.51)$$

und für GLS erhält man (siehe etwa SCHNEEWEISS, 1980, S. 55)

$$\hat{\beta}_2^{\text{GLS}} = \frac{A + (4\lambda(\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y}))^2 + A^2)^{1/2}}{2 \sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})} \quad (4.52)$$

mit $A = \sum_{t=1}^T (y_t - \bar{y})^2 - \lambda \sum_{t=1}^T (x_t - \bar{x})^2$ und $\lambda = \sigma_\varepsilon^2 / \sigma_v^2$. Für $E(v_t \varepsilon_t) = 0$ läuft die für GLS nötige Kenntnis (bis auf eine skalare multiplikative Konstante) der Kovarianzmatrix Σ aller Fehlerterme des Modells gerade auf die Kenntnis von λ hinaus.

Wie in Kapitel 3 geschieht der Vergleich der Schätzverfahren an Hand verschiedener Kriterien, mit Schwerpunkt auf den durch Monte Carlo Simulationen gewonnenen empirischen Mittleren Quadratischen Abweichungen ($MSE(\hat{\beta}_2^{\text{OLS}})$, $MSE(\hat{\beta}_2^{\text{GLS}})$, und $MSE(\hat{\beta}_2^{\text{CLS}})$). Für ε_t und v_t unabhängig und normalverteilt geben RICHARDSON und WU (1970) exakte Formeln für Verzerrung und Varianz von OLS, woraus sich für $MSE(\hat{\beta}_2^{\text{OLS}})$ ergibt (S. 742):

$$\begin{aligned} MSE(\hat{\beta}_2^{\text{OLS}})(a) &= E (\hat{\beta}_2^{\text{OLS}} - \beta_2)^2 \\ &= \beta_2^2 \left\{ \frac{T-3}{2} + \frac{1}{T-2} \left[\frac{\sigma_\varepsilon^2}{\beta_2^2 \sigma_v^2} - ((T-3)w + (T-5)\left(\frac{T}{2} - 1\right)) \right] \right. \\ &\quad \left. \cdot e^{-w} M\left(\frac{T}{2} - 1, \frac{T}{2}, w\right) \right\}. \end{aligned} \quad (4.53)$$

Dabei ist $w = \sum_{t=1}^T (z_t - \bar{z})^2 / 2\sigma_v^2$, und $M(, ,)$ steht für die in der Verteilungstheorie ökonomischer Schätzverfahren des öfteren benötigte Confluente Hypergeometrische Funktion (siehe SLATER, 1965).

Der Index (a) dient zur Unterscheidung von den aus Monte Carlo Simulationen gewonnenen empirischen Mittlere Quadratische Abweichungen, die im weiteren mit einem Index (e) bezeichnet werden.

Da für $M(\cdot, \cdot, \cdot)$ außer in Spezialfällen kein numerisch leicht zugänglicher geschlossener Ausdruck existiert, wurde $M(T/2 - 1, T/2, w)$ durch die Rekursion

$$e^{-w} M(T/2 - 1, T/2, w) = \frac{T-2}{2w} (1 - e^{-w} M(T/2 - 2, T/2 - 1, w)) \quad (4.54)$$

mit dem Startwert $e^{-w} M(0, 1, w) = e^{-w}$ berechnet (T gerade). Für ungerade T ist der Startwert $e^{-w} M(1/2, 3/2, w)$ nicht unmittelbar anzugeben.

(4.53) erlaubt eine Effizienzsteigerung der direkten Monte Carlo Schätzungen $MSE(\hat{\beta}_2^{GLS})(e)$ und $MSE(\hat{\beta}_2^{CLS})(e)$ durch Verwendung von Kontrollvariablen (siehe etwa HAMMERSLEY und HANDSCOMB, 1964, S. 59). Sei allgemein $\hat{\xi}$ eine Monte Carlo Schätzung für einen skalaren Parameter ξ , und $\hat{\theta}$ eine weitere Monte Carlo Statistik mit bekanntem Erwartungswert θ . Bei direkter Simulation (d.h. $\hat{\xi} = \frac{1}{N} \sum_{i=1}^N \xi_i$ ist das arithmetische Mittel von N in unabhängigen Wiederholungen gewonnenen Einzelwerten einer Pseudo-Zufallsvariablen mit Erwartungswert ξ) ist $\hat{\xi}$ eine erwartungstreue Schätzung für ξ . Jedoch ist $\hat{\xi}^{(c)} = \hat{\xi} - \hat{\theta} + \theta$ ebenfalls erwartungstreu, und hat für

$$\text{cov}(\hat{\xi}, \hat{\theta}) > \frac{1}{2} \text{var}(\hat{\theta}) \quad (4.55)$$

eine kleinere Varianz. $\hat{\theta} - \theta$ wird dabei Kontrollvariable genannt.

Die im weiteren verwendete Kontrollvariable ist $MSE(\hat{\beta}_2^{OLS})(a) - MSE(\hat{\beta}_2^{OLS})(e)$, die Differenz der mit (4.53) analytisch berechneten und der durch direkte Monte Carlo Simulation geschätzten mittleren qua-

dratischen Abweichung von OLS. $MSE(\hat{\beta}_2^{\text{GLS/CLS}})(c) := MSE(\hat{\beta}_2^{\text{GLS/CLS}})(e) - MSE(\hat{\beta}_2^{\text{OLS}})(e) + MSE(\hat{\beta}_2^{\text{OLS}})(a)$ ist damit die mit der Kontrollvariablen korrigierte Monte Carlo Schätzung für $MSE(\hat{\beta}_2^{\text{GLS/CLS}})$.

Ob oder ob nicht das einen Effizienzgewinn bedeutet, hängt von (4.55) ab. Diese Größen sind nicht bekannt, können aber aus den Simulationsergebnissen geschätzt werden. Wenn immer dadurch ein Effizienzgewinn angezeigt wird, erscheint in den folgenden Tabellen $MSE(\hat{\beta}_2^{\text{GLS/CLS}})(c)$, ansonsten $MSE(\hat{\beta}_2^{\text{GLS/CLS}})(e)$. Offenbar liefert die Kontrollvariable vor allem bei hoher positiver Korrelation der empirischen zweiten Momente einen Effizienzgewinn. Dieser erreichte in vielen Experimenten einen Faktor von mehr als 10, d.h. statt der standardmäßigen 200 Wiederholungen wären für die mit Kontrollvariablen erzielte Simulationsgenauigkeit bei direkter Simulation $10 \times 200 = 2000$ Durchgänge nötig gewesen.

Dies gilt mit einer wichtigen Einschränkung: Im vorliegenden Modell existiert weder die Varianz von $\hat{\beta}_2^{\text{GLS}}$ noch die Varianz von $\hat{\beta}_2^{\text{CLS}}$ (ANDERSON und SAWA, 1982; KETELLAPPER, 1983). Wie schon in Kapitel 3 betont, sind damit die empirischen mittleren quadratischen Abweichungen als Schätzungen für die (immer existierenden) entsprechenden Populationsparameter der bei der größtmöglichen Maschinenkonstanten gestutzten Verteilung zu interpretieren.

Das gilt immer und unabhängig davon, ob die Momente der ungestutzten Verteilung existieren oder nicht. Implizit wird daher im weiteren unterstellt, daß bei existierenden Momenten die durch die Stützung bewirkte Differenz für die Zwecke der vorliegenden Untersuchung vernachlässigbar ist (OLS), oder daß bei nicht existierenden Momenten der ungestutzten Verteilung die Momente der simulierten

Verteilung die Variabilität der ungestutzten Verteilung geeignet charakterisieren. Analoge Voraussetzungen liegen auch der oft geübten Praxis zugrunde, Momente von asymptotischen Verteilungen zur Beschreibung von endlichen-Stichproben Verteilungen mit nicht notwendig existierenden Momenten heranzuziehen. Dies entspricht auch der Vorgangsweise vergleichbarer Monte Carlo Studien, im vorliegenden Modellzusammenhang etwa KETELLAPPER (1982).

Tabellen (4.1) und (4.2) fassen eine Reihe von Experimenten zur Überprüfung der endlichen-Stichproben Relevanz von Satz (4.2) zusammen (der asymptotischen Äquivalenz von OLS, GLS und CLS bei Trend in allen fehlerbehafteten exogenen Variablen). Der besseren Vergleichbarkeit wegen werden für jede Parameterkonstellation nur die Quotienten der mittleren quadratischen Abweichungen angegeben (unter Verwendung des exakten Werts für OLS und des empirischen bzw durch Kontrollvariable korregierten Werts bei GLS und CLS). ε_t und v_t ($t=1, \dots, T$) sind normalverteilt, um (4.53) anwenden zu können. In allen Experimenten wurde $\sigma_\varepsilon^2 = 10$ gesetzt. $\lambda = \sigma_\varepsilon^2 / \sigma_v^2$ erhält die Werte 5, 2, 1 und 0.2, β_2 die Werte 0.5, 2 und 8 (das impliziert für σ_u^2 Werte zwischen 10.5 und 3210). T variiert von 20 über 50, 100 und 200 bis 400. Das ergibt insgesamt 60 Parameterkombinationen (= 60 Experimente). Diese Parameter wurden nach ihrer möglichen Bedeutung für die empirische Praxis ausgewählt, mit Ausnahme von $\lambda = .2$. Die dadurch implizierte fünfmal größere Varianz des Meßfehlers im Vergleich zur Varianz des Fehlers in der Gleichung dürfte in der ökonometrischen Praxis eher selten sein und wird hier nur mit der Absicht zugrundegelegt, das Verhalten der Schätzverfahren auch in Extremsituationen zu verfolgen. Pro Parameterkombination wurden 200 Wiederholungen durchgeführt.

TABELLE (4.1): Relative Mittlere Quadratische Abweichungen
 $(z_t = t(\text{mod. } 20))$

	β_2											
	.5				2				8			
λ :	5	2	1	.2	5	2	1	.2	5	2	1	.2
	a) MSE(OLS)/MSE(GLS)											
T												
20	.98	1.02	1.02	.17	1.24	1.35	1.70	1.82	1.45	1.61	2.37	2.45
50	1.04	1.15	1.58	3.30	2.05	4.15	5.97	9.11	3.25	5.80	7.76	8.43
100	1.16	2.03	3.13	5.76	3.43	7.31	12.60	23.53	5.03	10.79	18.02	25.38
200	1.31	2.67	6.63	15.50	6.66	15.38	25.21	37.32	9.31	22.45	34.13	34.34
400	2.02	5.41	14.72	30.97	9.78	24.63	50.35	90.59	19.89	34.35	85.30	>100
	b) MSE(OLS)/MSE(CLS)											
20	.99	.95	.80	.02	1.14	1.11	.77	.00	1.23	1.19	1.32	.03
50	1.04	1.15	1.41	.01	1.85	3.01	3.53	.20	2.82	3.81	4.86	.01
100	1.16	1.99	2.83	1.41	3.34	5.84	8.48	3.03	4.14	7.95	9.83	.90
200	1.27	2.69	6.69	6.71	6.28	11.71	15.21	5.94	8.45	18.37	19.19	5.39
400	2.01	5.14	14.41	12.51	9.14	20.50	40.17	19.88	18.60	28.12	47.73	22.24

TABELLE (4.2): Relative Mittlere Quadratische Abweichungen
 $(z_t = t)$

	β_2											
	.5				2				8			
λ :	5	2	1	.2	5	2	1	.2	5	2	1	.2
	a) MSE(OLS)/MSE(GLS)											
T												
20	.98	1.02	1.02	.17	1.24	1.35	1.70	1.82	1.45	1.61	2.37	2.45
50	1.07	1.04	1.42	3.22	1.24	1.30	2.56	6.42	1.56	1.77	2.36	6.51
100	1.02	1.09	1.15	3.23	1.12	1.65	2.08	6.75	1.19	1.62	2.20	4.14
200	1.00	1.02	1.25	2.17	1.06	1.23	1.39	4.13	1.12	1.43	1.76	3.60
400	1.00	1.00	1.03	1.89	1.00	1.01	1.06	2.17	1.01	1.01	1.06	2.44
	b) MSE(OLS)/MSE(CLS)											
20	.99	.95	.80	.02	1.14	1.11	.77	.00	1.23	1.19	1.32	.03
50	1.06	1.04	1.38	2.72	1.25	1.27	2.38	3.94	1.47	1.62	2.12	3.95
100	1.02	1.10	1.15	2.85	1.13	1.62	2.02	5.48	1.81	1.62	2.06	3.76
200	1.00	1.02	1.25	2.21	1.06	1.25	1.38	4.18	1.12	1.42	1.70	3.24
400	1.00	1.00	1.03	1.91	1.00	1.01	1.06	2.16	1.01	1.01	1.06	2.45

In Tabelle (4.1) ist $z_t = t \pmod{20}$, d.h. die ersten 20 Werte werden bei wachsendem T wiederholt. Das entspricht dem in der Standardasymptotik behandelten Fall exogener Variablen ohne Trend und ist zugleich Ausgangsbasis aller weiteren Vergleiche.

Wie die Tabelle zeigt, nimmt die relative Effizienz von OLS für jede gegebene Parameterkonstellation mit wachsendem T monoton ab. Das ist nach der asymptotischen Theorie zu erwarten. Ohne Trend gilt $E(\hat{\beta}_2^{\text{OLS}} - \beta_2) \rightarrow c \neq 0$, und wegen

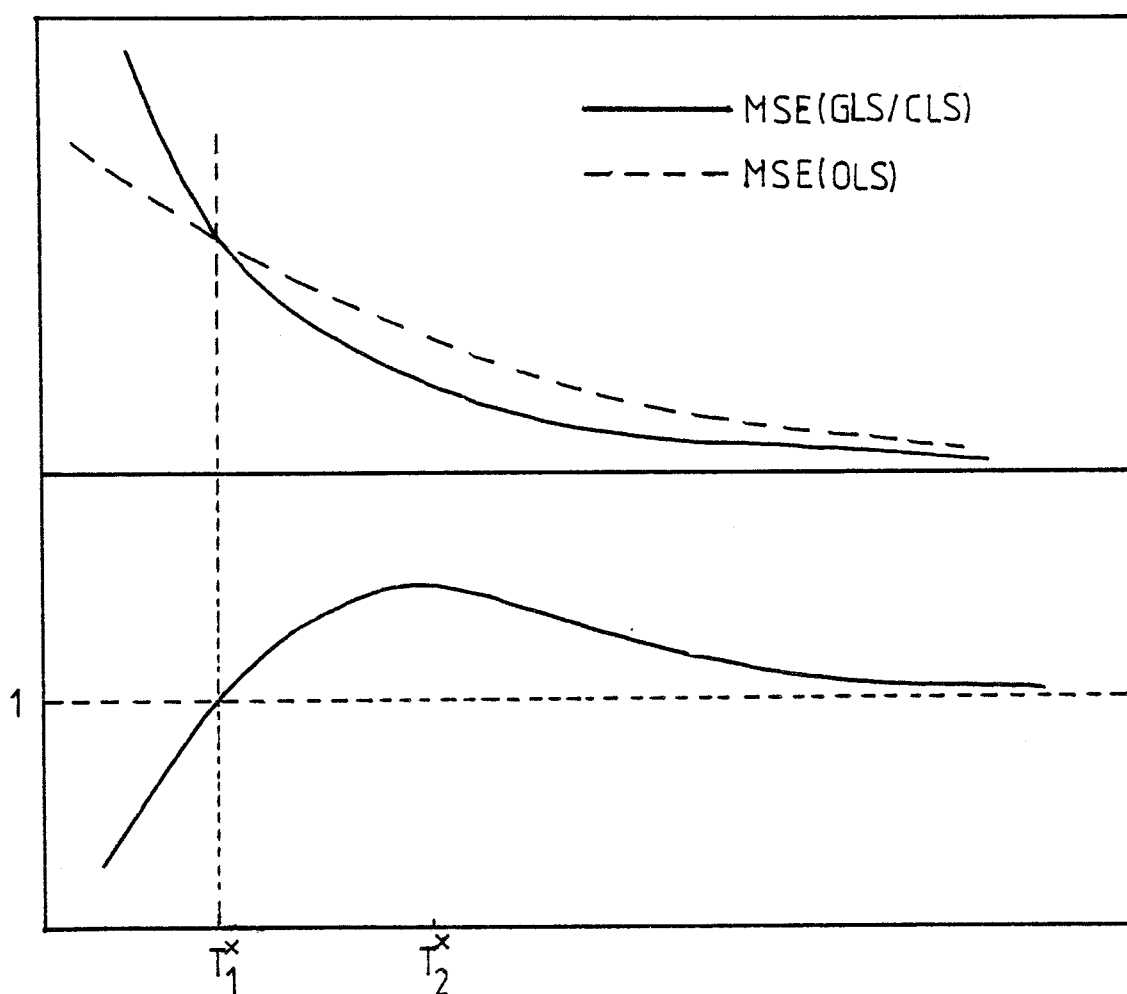
$$\text{MSE}(\hat{\beta}_2^{\text{OLS}}) = (E(\hat{\beta}_2^{\text{OLS}} - \beta_2))^2 + \text{var}(\hat{\beta}_2^{\text{OLS}}) \quad (4.56)$$

kann man daher im Standardfall kein Verschwinden (bei $T \rightarrow \infty$) von $\text{MSE}(\hat{\beta}_2^{\text{OLS}})$ erwarten, im Gegensatz zu den mittleren quadratischen Abweichungen der (gestutzten) Verteilungen von GLS und CLS. Damit gehen aber die jeweiligen Quotienten gegen ∞ .

Wenig überraschend ist auch die (ceteris paribus) mit wachsender Varianz der Meßfehler (fallendem λ) abnehmende Qualität von OLS. Nur bei kleinen Stichproben wird dies aufgehoben und teilweise überkompensiert durch die negativen Konsequenzen einer hoher Meßfehlervarianz für die Varianz von GLS und CLS. Wie etwa aus (4.51) unmittelbar zu sehen, bewirkt wachsendes σ_v^2 bei CLS ceteris paribus eine Verschiebung der Verteilung des Nenners zum Ursprung und damit eine wachsende Variabilität des Quotienten (zugleich liefert das einen intuitiven Hintergrund für die Nichtexistenz der Momente von CLS). Wie die Tabelle zeigt, führt dieser Effekt in kleinen Stichproben (im Sinn von $\sigma_z^2 \approx \sigma_v^2$) zu mitunter drastischen Aufblähungen der empirischen Varianz relativ zu OLS.

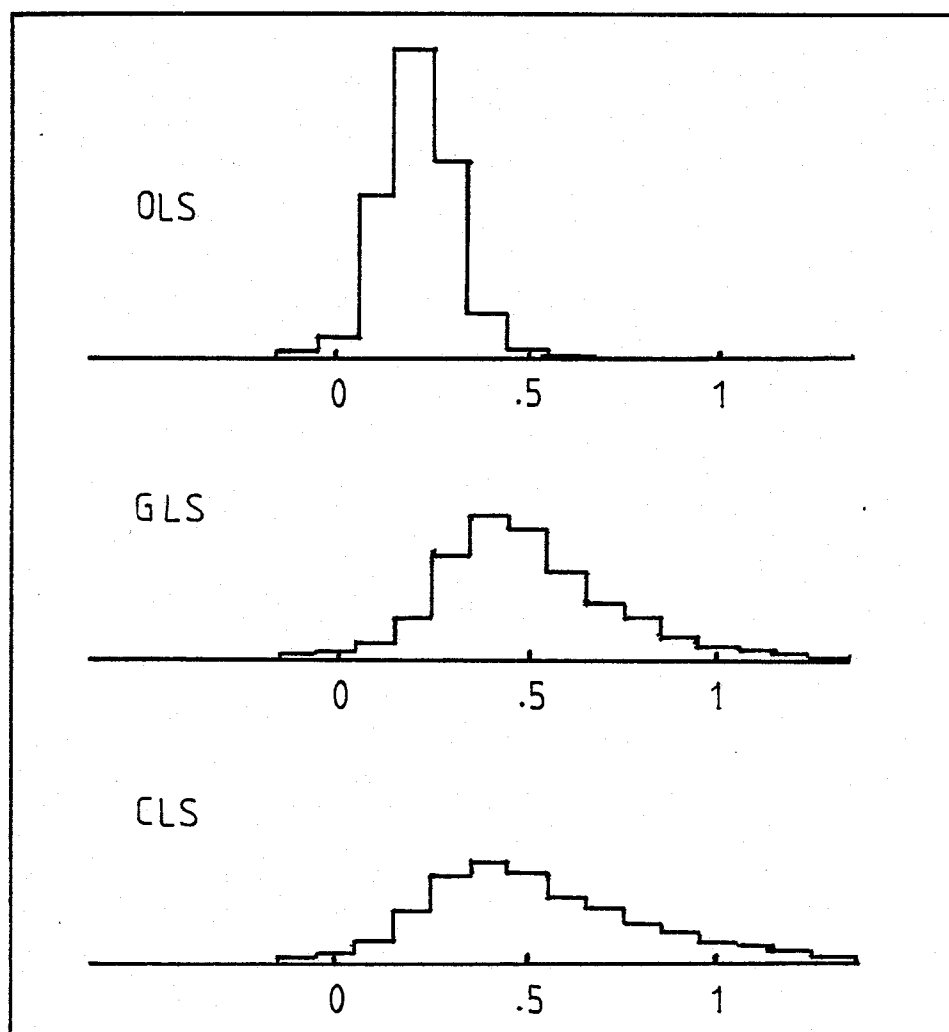
Tabelle (4.2) reproduziert zum Vergleich die analogen Simulationsergebnisse für $z_t = t$, d.h. einen linearen Trend in der exogenen Variablen. Für $T=20$ sind die exogenen Variablen und damit auch die relativen Effizienzen mit Tabelle (4.1) identisch. Insbesondere ist OLS für einige Parameterkonstellationen besser als GLS und CLS. Mit wachsendem T nimmt die relative Effizienz von OLS jedoch wie im Fall ohne Trend zunächst ab, ist für ein von den anderen Parametern abhängiges T_1^* gleich 1, erreicht bei T_2^* ein Minimum und strebt dann dem Grenzwert 1 zu. Wie schon in Kapitel 3 steigt die relative Effizienz von OLS daher keineswegs monoton. Schaubild (4.2) illustriert diesen Sachverhalt.

SCHAUBILD (4.2): Typischer Verlauf der relativen Effizienz von OLS bei Trend in z_t



Der vergleichsweise flache Verlauf der MSE-Kurve für OLS geht dabei auf die dominierende Rolle der quadrierten Verzerrung als Komponente der Mittleren Quadratischen Abweichung von OLS (nicht aber GLS und CLS) zurück. Diese verschwindet in den Experimenten mit Trend für kleine und wachsende T langsamer als die Varianz (und ohne Trend überhaupt nicht). Zur Illustration wurde das durch $\beta = .5$, $\lambda = 5$, $T = 20$ definierte Experiment mit 10 000 Durchgängen wiederholt. Schaubild (4.3) gibt die Histogramme der dabei erzielten Schätzwerte wieder. Man sieht deutlich die relativ große Verzerrung von OLS, die

SCHAUBILD (4.3): Histogramm von 10 000 Realisationen von $\hat{\beta}_2^{\text{OLS}}$, $\hat{\beta}_2^{\text{GLS}}$ und $\hat{\beta}_2^{\text{CLS}}$



aber für diese spezielle Parameterkonstellation durch die im Vergleich zu OLS ungleich höhere Varianz von GLS und CLS mehr als ausgeglichen wird. Die daraus resultierende mögliche MSE-Überlegenheit von OLS in kleinen Stichproben ist wohlbekannt und hat zu Vorschlägen geführt, OLS mit anderen Verfahren linear zu kombinieren, bei mit wachsendem T abnehmender Gewichtung von OLS, wie in FELDSTEIN (1974).

Ein weiteres Ergebnis dieser ersten Versuchsreihe, das insbesondere in Tabelle (4.2) zum Ausdruck kommt, ist die vergleichsweise schnelle Konvergenz gegen 1 der relativen Effizienz von GLS und CLS. Die Quotienten der Mittleren Quadratischen Abweichungen sind nicht direkt in der Tabelle angegeben, können aber unmittelbar aus $MSE(OLS)/MSE(GLS)$ und $MSE(OLS)/MSE(CLS)$ gewonnen werden. Dies deutet darauf hin, daß die Differenz zweier auch ohne Trend konsistenter Schätzverfahren bei Trend schneller verschwindet als die Differenz zwischen diesen Verfahren auf der einen und OLS auf der anderen Seite, eine etwa durch SCHNEEWEISS (1982) nahegelegte Möglichkeit.

In weiteren Versuchsreihen wurde die Abhängigkeit dieser Simulationsergebnisse von den zugrundegelegten exogenen Variablen überprüft. Tabelle (4.3) und Tabelle (4.4) geben die komprimierten Ergebnisse wieder. In Tabelle (4.3) wurde der Beveridge-Weizenpreis-Index für z_t verwendet, als Beispiel einer real beobachteten ökonomischen Zeitreihe mit Trend. Für $t=1, \dots, 370$ (1500 - 1869) sind die Werte aus ANDERSON (1970, S. 623-626) entnommen, mit z_1, \dots, z_{110} wie in Schaubild (4.1). Die bei $T=400$ fehlenden letzten 30 Werte sind Wiederholungen von z_{340}, \dots, z_{369} . In Tabelle (4.4) wurde $z_t = t^2$ gesetzt.

Die Zahlen in Tabelle (4.3) bestätigen die Simulationsergebnisse für $z_t = t$: Für kleine Stichproben (im Sinn von σ_z^2 "klein" im Vergleich zu σ_v^2) ist OLS dank kleinerer Varianz gegenüber GLS und besonders gegen-

über CLS überlegen, aber diese Überlegenheit verkehrt sich mit wachsendem T zunächst ins Gegenteil, bis dann die asymptotische Gleichwertigkeit der Verfahren durchzuschlagen beginnt. Aus den Ergebnissen für $z_t = t^2$ schließlich wird deutlich, daß die asymptotische Gleichwertigkeit bei stärkerem als linearem Trend schon sehr viel früher wirksam wird.

Eine letzte Reihe von Versuchen gilt der Überprüfung der asymptotischen Normalverteilung von OLS. Bei Trend in z_t gilt wegen Satz (4.2), in Verbindung mit (4.47), (4.48) und Satz (4.3), daß

$$\left(\sum_{t=1}^T (x_t - \bar{x})^2 / s^2\right)^{1/2} (\hat{\beta}_2^{\text{OLS}} - \beta_2) \xrightarrow{D} N(0,1) . \quad (4.57)$$

Mit $z_t = t$, und für $\beta_2=2$, $\sigma_v^2=5$, $\sigma_\varepsilon^2=10$ (d.h. $\lambda = 2$ und $\sigma_u^2=30$) wurden die entsprechenden Experimente aus Tabelle (4.2) für $T=20$, 100 und 400 mit je 10 000 Durchgängen wiederholt. Schaubild (4.4) gibt die Histogramme der dabei gewonnenen Realisationen von (4.47) wieder.

Man sieht, daß zumindest für die hier gewählte Parameterkonstellation die Normalverteilungsapproximation mit wachsendem T zwar besser wird, aber auch für $T=400$ noch keinesfalls als zufriedenstellend anzusehen ist. Zur Konkretisierung stellt Tabelle (4.5) einige nominelle und tatsächliche Signifikanzniveaus des auf (4.57) basierenden Tests von $H_0: \beta_2=2$ gegenüber. Die ungewohnten nominellen Signifikanzniveaus sind dabei eine Folge der für Schaubild (4.4) gewählten Intervallgrenzen.

Wie die Tabelle zeigt, ist die wahre Wahrscheinlichkeit für einen Fehler erster Art bei linksseitigen Tests größer und bei rechtsseitigen Tests kleiner als das nominelle Signifikanzniveau, ein auch unmittelbar aus Schaubild (4.4) abzulesendes Resultat. Diese Diffe-

SCHAUBILD (4.4): Histogramme von je 10 000 Realisationen des normierten Schätzfehlers von OLS ($\beta_2=2$, $\sigma_v^2=5$, $\sigma_\varepsilon^2=10$)

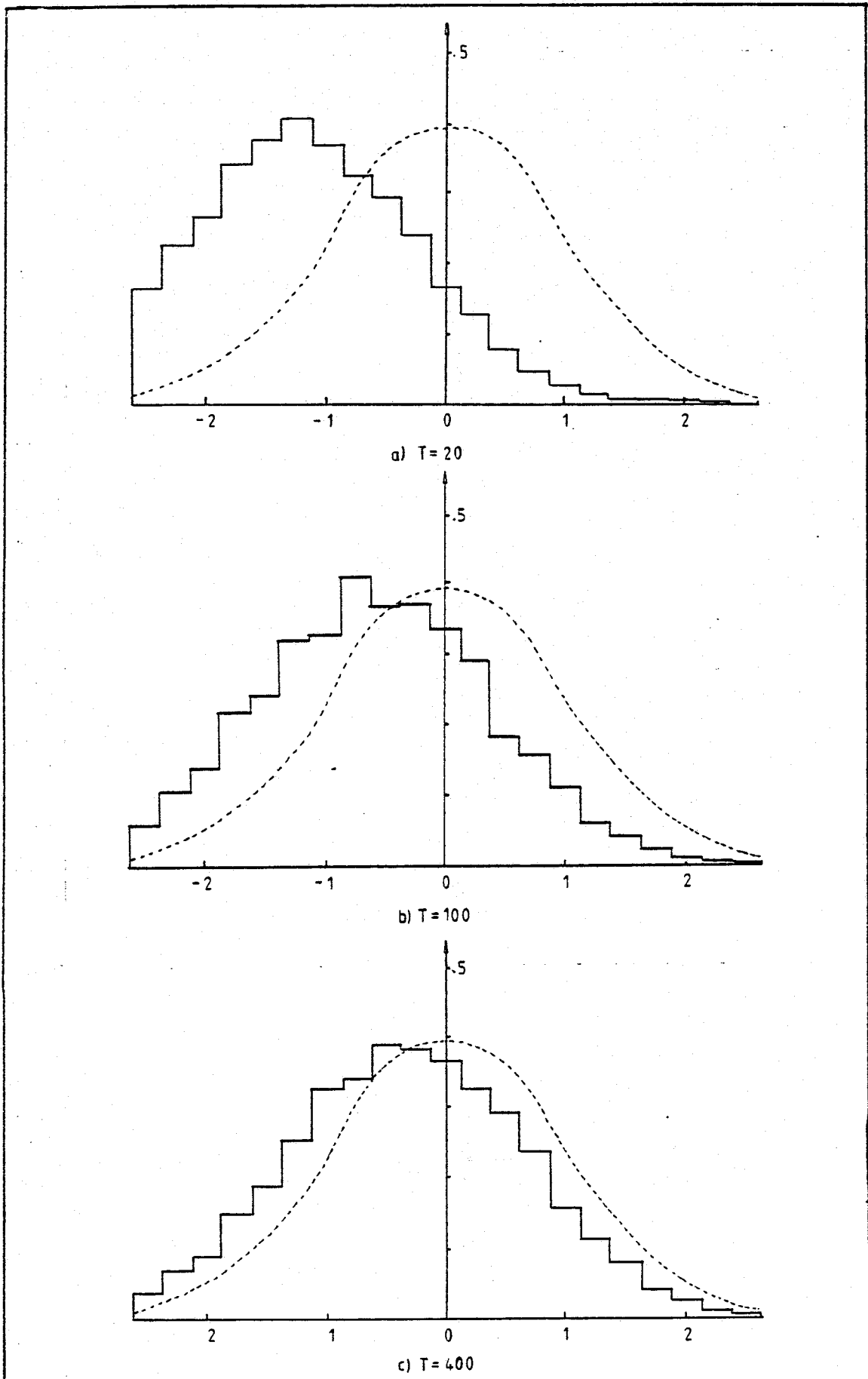


TABELLE (4.5): Vergleich von nominellem und tatsächlichen
Signifikanzniveau beim Test von $H_0: \beta_2=2$

	H_1					
	$\beta_2 \leq 2$		$\beta_2 \geq 2$		$\beta_2 \neq 2$	
$\alpha(\text{nominell}):$	5.2%	.88%	5.2%	.88%	10.4%	1.76%
T=20	36.4%	15.18%	.37%	.06%	36.8%	15.24%
$\alpha(\text{real}):$ T=100	16.3%	4.08%	1.28%	.19%	17.5%	4.27%
T=400	9.9%	2.06%	2.52%	.36%	12.5%	2.42%

renzen nehmen mit wachsendem T ab und gleichen sich bei zweiseitigen Tests teilweise aus. Ob dies auch auf multiple Regressionen übertragbar ist, in denen die Richtung der Verzerrung von OLS nicht sofort wie im bivariaten Fall feststeht, sei hier dahingestellt. Für die ökonometrische Praxis mag auch der Hinweis von Interesse sein, daß derartige Abweichungen zwischen nominellen und tatsächlichen Signifikanzniveaus beim Standardsignifikanztest von $H_0: \beta_2=0$ wegen der dann unter H_0 geltenden Unabhängigkeit von beobachteten Regressoren und Störgrößen nicht auftreten können.

d) Diskussion

Die gegenwärtige Renaissance des Fehler-in-den-Variablen Modells dokumentiert sich naturgemäß in einer immer umfangreicheren Spezialliteratur. Nützliche Übersichten liefern MADANSKI (1959), GRILICHES (1974), SCHNEEWEISS (1980), und insbesondere AIGNER et al. (1983). SCHNEEWEISS und MITTAG (1984) geben eine detaillierte Lehrbuchdarstellung des Gebiets, und auf die durch das Fehler-in-den-Variablen Modell ermöglichten Querverbindungen zwischen Psychometrie und Ökonometrie machen unter anderem JÖRESKOG und GOLDBERGER (1975), AIGNER und GOLDBERGER (1977) oder de LEEUW et al. (1983) aufmerksam.

Ähnlich wie in Abschnitt 4 b) weist schon SCHNEEWEISS (1982) auf die durch Trend bewirkte Konsistenz von OLS trotz Fehlern in den Variablen hin, allerdings im Rahmen eines einfachen bivariaten Modells. Zusätzlich ist seine Betrachtung, wie auch in KRÄMER (1983 b), auf linearen Trend beschränkt. Die asymptotische Normalität von CLS unter den Bedingungen (SCH), die wegen Satz (2.1) geringfügig restriktiver sind als die hier verwendeten Bedingungen (K) und (GR), zeigt auch schon SCHNEEWEISS (1976).

Im Einklang mit der Literatur wurden in diesem Kapitel wie auch in Kapitel 3 die Störgrößen- und Fehlervarianzen auch bei Trend in den exogenen Variablen konstant gehalten. Das impliziert im Trendfall einen mit wachsenden Stichproben verschwindenden relativen Beitrag der Meßfehler zur empirischen Varianz der beobachteten Regressoren, wie auch einen verschwindenden relativen Beitrag der Störgrößenvarianz in der Gleichung zur empirischen Varianz der endogenen Variablen, und beinhaltet damit eine in konkreten Anwendungen jeweils neu zu recht-

fertigende Voraussetzung.

Derartige durch verschwindende relative Beiträge der Fehler charakterisierte Modellsituationen beleuchtet auch Wolds sogenanntes Proximity Theorem (WOLD und FAXER, 1957), oder die von KADANE (1971) initiierte klein- σ -Asymptotik. Im Gegensatz zu diesen Ansätzen sind aber hier die Kreuzproduktmatrizen der Regressoren nicht durch einen skalaren Faktor normalisierbar.

Die strukturelle Variante des Fehler-in-den-Variablen Modells, mit stochastischen und identisch verteilten z_t 's, läßt per definitionem keinen Trend zu (Existenz der zweiten Momente vorausgesetzt) und bleibt daher in diesem Kapitel außer Betracht. Jedoch ist die Beschränkung auf nichtstochastische exogene Variable unnötig restriktiv, und lassen sich analoge Ergebnisse auch für stochastische exogene Variable mit deterministischem Trend in den Momenten ableiten, wie in SCHNEEWEISS (1976). Geeignete Kandidaten wären hier etwa die schon in Kapitel 3 als Regressoren untersuchten ARIMA-Prozesse.

5. TREND IN SIMULTANEN GLEICHUNGEN

a) Modell und Schätzverfahren

Im Mittelpunkt des vorliegenden Kapitels steht das vollständige, lineare, simultane Gleichungssystem

$$YB + Z\Gamma = U \quad (5.1)$$

Dabei ist Y eine $T \times G$ -Matrix von Beobachtungen der G endogenen Variablen des Systems und Z eine nichtstochastische $T \times K$ -Matrix von Beobachtungen der K exogenen Variablen. B ($G \times G$, regulär) und Γ ($G \times K$) sind die zu schätzenden Regressionskoeffizienten des Modells (im weiteren auch Strukturparameter genannt). U ($T \times K$) ist eine Matrix stochastischer Störgrößen, mit Erwartungswert 0 und unabhängigen, identisch verteilten Zeilen mit endlicher Varianzmatrix Σ .

Die zu schätzende Gleichung aus (5.1) sei

$$y^{(1)} = Y^{(1)}\beta^{(1)} + Z^{(1)}\gamma^{(1)} + u^{(1)} \quad (5.2)$$

Dabei enthalten $y^{(1)}$ ($T \times 1$) und $Y^{(1)}$ ($T \times g_1$) die Beobachtungen der $g_1 + 1$ endogenen Variablen der Gleichung, $Z^{(1)}$ ($T \times K_1$; Rang K_1) ist die Beobachtungsmatrix der K_1 in der Gleichung vertretenen exogenen Modellvariablen, und $u^{(1)}$ ($T \times 1$) ist der Störgrößenvektor der Gleichung.

σ_u^2 bezeichne die Varianz der Komponenten von $u^{(1)}$. Weiterhin sei das Rangkriterium für die Identifikation von (5.2) erfüllt, d.h. insbesondere $g_1 + K_1 \leq K$.

Untersuchungsgegenstand des vorliegenden Kapitels sind wie in den vorhergehenden Abschnitten die asymptotischen Eigenschaften einiger Schätzverfahren, und dabei besonders von OLS, für die unbekanntem Strukturparameter des Modells. Die konventionelle große-Stichproben-Asymptotik geht dabei von der Annahme aus, daß

$$\lim_{T \rightarrow \infty} T^{-1} Z'Z = Q \quad (\text{regulär}) \quad (5.3)$$

(siehe THEIL, 1971, S. 485; SCHÖNFELD, 1971, S. 161; JOHNSTON, 1972, S. 351). (5.3) entspricht der Annahme (4.6) im Fehler-in-Variablen Modell oder der Annahme (2.4) bei linearer Standardregression, und genau wie dort ist durch (5.3) ein Trend in den exogenen Variablen des Modells explizit ausgeschlossen.

Mit

$$X = (Y^{(1)} : Z^{(1)}) \quad , \quad \beta = (\beta^{(1)}, \gamma^{(1)})' \quad (5.4)$$

ist $\hat{\beta}^{OLS} = (X'X)^{-1} X'y^{(1)}$ die OLS-Schätzung der Strukturparameter in (5.2). Unter (5.3) ist $\hat{\beta}^{OLS}$ im allgemeinen nicht konsistent, eines der bekanntesten Resultate der ökonometrischen Methodologie.

Zur Illustration betrachte man HAAVELMOS (1947) Ableitung der asymptotischen Verzerrung der OLS-Schätzung der marginalen Konsumquote in einem einfachen Keynes'schen Modell, mit Konsumfunktion

$$C_t = \beta_1 + \beta_2 Y_t + u_t \quad (t=1,2,\dots) \quad (5.5)$$

und Bilanzgleichung

$$Y_t = C_t + A_t \quad (5.6)$$

Dabei steht C_t für die Konsumausgaben, Y_t für das Volkseinkommen, und A_t für die autonomen Ausgaben (Investition und Staatsausgaben) einer Periode t . Trotz abweichender Notation ist das System (5.5) - (5.6) offenbar ein Spezialfall von (5.1), mit $G = K = 2$, und einer Spalte von U identisch Null. Die endogenen Variablen des Systems sind C und Y , die exogenen die Konstante und A . Zur besseren Vergleichbarkeit mit Haavelmo und der einschlägigen Diskussion in der Lehrbuchliteratur (siehe etwa SCHÖNFELD, 1971, S. 225, oder JOHNSTON, 1972, S. 342-344) wird aber hier die Originalschreibweise soweit möglich beibehalten.

In Haavelmos Notation ist damit die OLS-Schätzung für die marginale Konsumquote gegeben durch

$$\hat{\beta}_2^{\text{OLS}} = m_{cy} / m_{yy} \quad (5.7)$$

mit

$$m_{cy} = T^{-1} \sum_{t=1}^T (C_t - \bar{C})(Y_t - \bar{Y}) ; \quad m_{yy} = T^{-1} \sum_{t=1}^T (Y_t - \bar{Y})^2 . \quad (5.8)$$

Die asymptotische Verzerrung von OLS folgt dann unmittelbar aus der reduzierten Form des Systems:

$$C_t = \beta_1(1-\beta_2)^{-1} + \beta_2(1-\beta_2)^{-1}A_t + (1-\beta_2)^{-1}u_t \quad (5.9)$$

$$Y_t = \beta_1(1-\beta_2)^{-1} + (1-\beta_2)^{-1}A_t + (1-\beta_2)^{-1}u_t . \quad (5.10)$$

Mit m_{au} , m_{aa} und m_{uu} analog zu (5.8), liefern (5.9) und (5.10)

$$\hat{\beta}_2^{\text{OLS}} - \beta_2 = \frac{(1-\beta_2)(m_{au} + m_{uu})}{m_{aa} + m_{uu} + 2m_{au}} . \quad (5.11)$$

Die Annahme (5.3) ist aber im vorliegenden Beispiel äquivalent zu

$$0 < \lim_{T \rightarrow \infty} m_{aa} = \bar{m}_{aa} < \infty . \quad (5.12)$$

Bei konstanter und positiver Varianz σ^2 der u_t 's liefert dies sofort

$$m_{uu} \xrightarrow{P} \sigma^2, \quad m_{au} \xrightarrow{P} 0, \quad \text{und damit wegen (5.11)}$$

$$\text{plim}_{T \rightarrow \infty} (\hat{\beta}_2^{\text{OLS}} - \beta_2) = (1 - \beta_2) \frac{\sigma^2}{\bar{m}_{aa} + \sigma^2} \neq 0 \quad (5.13)$$

(außer für $\beta_2 = 1$, was aber sachlogisch ausgeschlossen ist). Damit ist OLS unter (5.3) nicht konsistent.

Zum Vergleich betrachte man nun den durch (5.3) ausgeschlossenen Fall trendbehafteter autonomer Ausgaben, wobei der Einfachheit halber $A_t = t$ ($t=1,2,\dots$) gesetzt sei. Jetzt gilt

$$\begin{aligned} \lim_{T \rightarrow \infty} T^{-1} m_{au} &= \lim_{T \rightarrow \infty} m_{uu} = 0, \\ \lim_{T \rightarrow \infty} T^{-2} m_{aa} &= 1/12 > 0, \end{aligned} \quad (5.14)$$

d.h. $\hat{\beta}_2^{\text{OLS}} - \beta_2 \xrightarrow{P} 0$. Aus (5.14) folgt sogar das weit stärkere Resultat $T(\hat{\beta}_2^{\text{OLS}} - \beta_2) \xrightarrow{P} 0$, d.h. OLS ist jetzt konsistent.

Obwohl vielleicht zunächst überraschend, wird dieses Ergebnis schon durch (5.13) nahegelegt. Dort zeigt sich, daß die asymptotische Verzerrung von OLS ceteris paribus mit wachsendem \bar{m}_{aa} (d.h. wachsender relativer Variabilität der exogenen Variablen) abnimmt. Das legt analog zu dem in 4 a) diskutierten Beispiel die Vermutung nahe, daß diese asymptotische Verzerrung für trendbehaftete exogene Variable, d.h. $m_{aa} \rightarrow \infty$ ganz verschwindet.

Als Konkurrenz zu OLS werden im weiteren vor allem k-Klasse Schätzungen untersucht. Mit

$$X^{(k)} = (Y^{(1)} - k\hat{V}^{(1)} : Z^{(1)}) , \quad (5.15)$$

$$\hat{V}^{(1)} = Y^{(1)} - Z(Z'Z)^{-1}Z'Y^{(1)} \quad (5.16)$$

ist eine k-Klasse Schätzung für die Strukturparameter in (5.2) gegeben durch

$$\hat{\beta}^{(k)} = (X^{(k)'}, X^{(k)})^{-1} X^{(k)'} Y^{(1)} . \quad (5.17)$$

Offenbar liefert $k=0$ gerade OLS und $k=1$ die Two-Stage-Least-Squares Schätzung (2SLS) für β . Auch Limited-Information-Maximum-Likelihood Schätzungen (LIML) lassen sich mit stochastischem k als k-Klasse Schätzungen auffassen. Falls

$$\text{plim}_{T \rightarrow \infty} k(T) = 1 , \quad (5.18)$$

sind diese Verfahren unter (5.3) konsistent (siehe etwa THEIL, 1971, Kap. 10, oder SCHÖNFELD, 1971, Kap. 17). Insbesondere wird OLS durch (5.18) bei Geltung von (5.3) aus der Klasse der konsistenten Schätzungen ausgeschlossen.

Wie schon das einführende Beispiel gezeigt hat, ändert sich das jedoch bei Trend in den exogenen Modellvariablen. Im weiteren wird dabei im Einklang mit der bisherigen Vorgangsweise Trend in der i 'ten Spalte von Z definiert durch $\|z_{\cdot i}\| / T^{1/2} = (\sum_{t=1}^T z_{ti}^2)^{1/2} / T^{1/2} \rightarrow \infty$, und für das Grenzverhalten von $Z'Z$ der Bedingungskomplex $(K) = (GR)$ unterstellt. Im vorliegenden Zusammenhang bedeutet das

$$\max_{1 \leq t \leq T} z_{ti}^2 / \sum_{t=1}^T z_{ti}^2 \longrightarrow 0 \quad (i=1, \dots, K) \quad (5.19)$$

und

$$D_T^{-1} Z' Z D_T^{-1} \longrightarrow Q \text{ (regulär)} \quad (5.20)$$

mit $D_T = \text{diag}(\|z_{\cdot 1}\|, \dots, \|z_{\cdot K}\|)$.

Wie schon weiter oben in Kapitel 4 betont, beinhaltet (5.20) bei mehr als einer trendbehafteten exogenen Variablen und für gewisse Trendverläufe (etwa linearer Trend, überlagert von Abweichungen) Restriktionen zu den Abweichungen vom Trend. Für eine Diskussion der Realitätsnähe der obigen Annahmen sei daher auch hier wieder auf den Anhang a der vorliegenden Arbeit verwiesen.

b) Konsistenz und Effizienz von OLS

Sei $X = (Y^{(1)} : Z^{(1)})$ wie in (5.4) die Matrix der rechtsseitigen Variablen aus (5.2). Da keine Verwechslungsgefahr besteht, entfällt hier ein separater Index zur Kennzeichnung der Gleichung. $Z^{(2)}$ enthalte die nicht in (5.2) vorkommenden Spalten von Z , d.h. $Z = (Z^{(1)} : Z^{(2)})$. $Y^{(1)} = Z \Pi^{(1)} + V^{(1)}$ sei die reduzierte Form von $Y^{(1)}$. Mit einer der Zerlegung von Z entsprechenden Zerlegung von $\Pi^{(1)}$ und $\tilde{V} = (V^{(1)} : 0)$ läßt sich X dann schreiben als $X = Z C + \tilde{V}$ mit

$$C = \begin{bmatrix} \Pi^{(1,1)} & I \\ \Pi^{(1,2)} & 0 \end{bmatrix}. \quad (5.21)$$

Per Annahme ist das Rangkriterium für Identifizierbarkeit von (5.2) erfüllt, d.h. C hat vollen Spaltenrang.

Die Normalisierung von $\hat{\beta}^{\text{OLS}} - \beta$ geschieht im weiteren durch

$$W = (C'Z'ZC)^{1/2} \quad (5.22)$$

Analog zu Kapitel 4 könnte man zunächst versuchen, die Diagonalmatrix \tilde{D} der Diagonalelemente von $C'Z'ZC$ oder die entsprechenden Elemente aus $X'X$ für diesen Zweck heranzuziehen, aber wie man leicht sieht ist $\lim \tilde{D}^{-1}C'Z'ZC\tilde{D}^{-1}$ nicht notwendig regulär, was eine Beweisführung wie in Kapitel 4 unmöglich macht.

In Fortführung der bisherigen Notation seien trendfreie Variable im weiteren durch ein Superskript (a), trendbehaftete Variable durch ein Superskript (b) gekennzeichnet. Damit läßt sich die Zerlegung $Z = (Z^{(1)}:Z^{(2)})$ verfeinern zu

$$Z = (Z^{(1a)}:Z^{(1b)}:Z^{(2a)}:Z^{(2b)}) \quad , \quad (5.23)$$

und C läßt sich in entsprechender Zerlegung schreiben als

$$C = \begin{pmatrix} \Pi^{(1,1a)} & I & 0 \\ \Pi^{(1,1b)} & 0 & I \\ \Pi^{(1,2a)} & 0 & 0 \\ \Pi^{(1,2b)} & 0 & 0 \end{pmatrix} \quad (5.24)$$

Von besonderer Bedeutung in der Zerlegung von C ist $\Pi^{(1,2b)}$, die Koeffizientenmatrix der endogenen erklärenden Variablen der Gleichung mit den aus der Gleichung ausgeschlossenen exogenen Variablen.

LEMMA (5.1): Hat $\Pi^{(1,2b)}$ vollen Spaltenrang, so gilt

$$\text{plim}_{T \rightarrow \infty} W(X'X)^{-1}W = \lim_{T \rightarrow \infty} W(C'Z'ZC)^{-1}W = I \quad (5.25)$$

BEWEIS: Per Definition von W gilt $W(C'Z'ZC)^{-1}W = I$

und damit das zweite Gleichheitszeichen in (5.25). Ferner hat man

$$X'X = C'Z'ZC + C'Z'\tilde{V} + \tilde{V}'ZC + \tilde{V}'\tilde{V} \quad (5.26)$$

und daher

$$W^{-1}X'XW^{-1} = I + W^{-1}C'Z'\tilde{V}W^{-1} + W^{-1}\tilde{V}'ZCW^{-1} + W^{-1}\tilde{V}'\tilde{V}W^{-1} \quad (5.27)$$

Wegen $W^{-1}C'Z'\tilde{V} = O_p(1)$ und $W^{-1} \rightarrow 0$ verschwinden aber der zweite und dritte Summand in (5.27) für $T \rightarrow \infty$, und es bleibt nur noch

$$W^{-1}\tilde{V}'\tilde{V}W^{-1} \xrightarrow{p} 0 \quad (5.28)$$

zu zeigen.

Dazu sei

$$g(T) = \min_{z_{\cdot i}} \text{mit Trend } \|z_{\cdot i}\| \quad (5.29)$$

Aus der Definition von Trend folgt $g(T) = o(T^{1/2})$, und im weiteren wird gezeigt, daß

$$\begin{bmatrix} g(T)I & 0 \\ 0 & D^{(1)} \end{bmatrix}^{1/2} W^{-1} \begin{bmatrix} g(T)I & 0 \\ 0 & D^{(1)} \end{bmatrix}^{1/2} = O(1) \quad (5.30)$$

mit $g(T)I$ passend zur linken oberen Teilmatrix $V^{(1)}, V^{(1)}$ von $\tilde{V}'\tilde{V}$ und $D^{(1)}$ aus der zu $Z = (Z^{(1)}:Z^{(2)})$ konformen Zerlegung von D . Da die übrigen Teilmatrizen von $\tilde{V}'\tilde{V}$ identisch Null sind, impliziert (5.30) unmittelbar (5.28).

Zum Beweis von (5.30) betrachte man

$$\begin{aligned}
 M &= \begin{bmatrix} g(T)^{-1}I & 0 \\ 0 & D^{(1)-1} \end{bmatrix} C'Z'ZC \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} \\
 &= \begin{bmatrix} I & 0 \\ 0 & g(T)D^{(1)-1} \end{bmatrix} C'G^{-1}Z'ZG^{-1}C \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} \quad (5.31)
 \end{aligned}$$

mit $G = g(T) I_{g_1+K_1}$. Sei $\lambda_{\min}(M)$ der kleinste Eigenwert von M , und M^* definiert durch

$$M^* = \begin{bmatrix} I & 0 \\ 0 & g(T)D^{(1)-1} \end{bmatrix} C'D^{*-1}Z'ZD^{*-1}C \begin{bmatrix} \cdot \\ \cdot \end{bmatrix} \quad (5.32)$$

mit $D^* = \text{diag}(g(T)I, D^{(1b)}, g(T)I, D^{(2b)})$. D^* entsteht aus der in (5.20) eingeführten Diagonalmatrix D durch Einsetzen von $g(T)I$ für $D^{(1a)}$ und $D^{(2a)}$, d.h. alle Diagonalelemente von D^{*-1} sind kleiner oder gleich den entsprechenden Diagonalelementen von G^{-1} , mit der Konsequenz, daß

$$\lambda_{\min}(M) \geq \lambda_{\min}(M^*) \quad (5.33)$$

Wie man weiter leicht sieht, gilt

$$\lim_{T \rightarrow \infty} M^* = C^*QC^* \quad (5.34)$$

mit

$$C^* = \begin{bmatrix} 0 & I & 0 \\ \Pi^{(1,1b)} & 0 & I \\ 0 & 0 & 0 \\ \Pi^{(1,2b)} & 0 & 0 \end{bmatrix} \quad (5.35)$$

Die Zerlegung von C^* ist dabei analog zu (5.24). Ex Annahme hat $\Pi^{(1,2b)}$ und damit C^* vollen Spaltenrang, d.h. C^*QC^* ist regulär.

Daraus folgt sofort

$$\lambda_{\max}(M^{-1}) \leq \lambda_{\max}(M^{*-1}) = o(1), \quad (5.36)$$

damit aber auch (5.30) und das Lemma.

q.e.d.

SATZ(5.2): Hat $\Pi^{(1,2b)}$ vollen Spaltenrang, so gilt

$$\text{plim}_{T \rightarrow \infty} \hat{\beta}^{\text{OLS}} - \beta = 0, \quad (5.37)$$

d.h. OLS ist konsistent.

BEWEIS: Die Behauptung folgt unmittelbar aus Lemma (5.1). Man hat

$$\begin{aligned} \hat{\beta}^{\text{OLS}} - \beta &= (X'X)^{-1}X'u^{(1)} \\ &= W^{-1}\{W(X'X)^{-1}W\}\{W^{-1}X'u^{(1)}\}. \end{aligned} \quad (5.38)$$

Wegen Lemma (5.1) sind die Ausdrücke in geschweiften Klammern stochastisch beschränkt, und der Satz folgt sofort aus $W^{-1} \rightarrow 0$. q.e.d.

Damit $\Pi^{(1,2b)}$ vollen Spaltenrang besitzt, darf die Zahl der Zeilen die Zahl der Spalten nicht unterschreiten. Damit folgt als notwendige Bedingung für die Erfüllung des Konsistenzkriteriums für OLS: Die Zahl der aus der Gleichung ausgeschlossenen trendbehafteten exogenen Variablen darf nicht kleiner sein als die Zahl der in der Gleichung enthaltenen rechtsseitigen endogenen Variablen.

Analog zu Kapitel 4 erscheint es weiter sinnvoll zu fragen,

inwieweit die hier angegebene hinreichende Konsistenzbedingung für OLS selbst wiederum auch notwendig ist. Das läßt sich in der Tat für den Spezialfall eines linearen Trends relativ leicht nachweisen.

SATZ (5.3): Es sei $Z'Z = O(T^3)$, und $\Pi^{(1,2b)}$ habe keinen vollen Spaltenrang. Dann ist OLS (bis auf eine Parametermenge vom Lebesgue-Maß 0) nicht konsistent.

BEWEIS: Bei linearem Trend ist (5.20) äquivalent zu

$$\begin{bmatrix} T^{-1}Z^{(a)}, Z^{(a)} & T^{-2}Z^{(a)}, Z^{(b)} \\ T^{-2}Z^{(b)}, Z^{(a)} & T^{-3}Z^{(b)}, Z^{(b)} \end{bmatrix} \rightarrow \tilde{Q} \quad (5.39)$$

mit einer endlichen und regulären Grenzmatrix \tilde{Q} , und $\hat{\beta}^{OLS} - \beta$ gegeben als die Lösung des folgenden modifizierten Normalgleichungssystems:

$$\begin{bmatrix} T^{-1}X^{(a)}, X^{(a)} & T^{-2}X^{(a)}, X^{(b)} \\ T^{-2}X^{(b)}, X^{(a)} & T^{-3}X^{(b)}, X^{(b)} \end{bmatrix} \times \begin{bmatrix} \hat{\beta}^{OLS(a)} - \beta^{(a)} \\ T(\hat{\beta}^{OLS(b)} - \beta^{(b)}) \end{bmatrix} \\ = \begin{bmatrix} T^{-1}X^{(a)}, u^{(1)} \\ T^{-2}X^{(b)}, u^{(1)} \end{bmatrix} \quad (5.40)$$

Dabei ist $X^{(a)} = Z^{(1a)}$, $X^{(b)} = (Z^{(1b)} : Y^{(1)})$, und $X = (X^{(a)} : X^{(b)})$ die Zerlegung von X in trendfreie und (möglicherweise) trendbehaftete Komponenten. $\beta = (\beta^{(a)}, \beta^{(b)})'$ ist die konforme Zerlegung von β . Falls $\Pi^{(1,2b)}$ vollen Spaltenrang besitzt, sieht man leicht, daß der erste Ausdruck in (5.40) gegen eine reguläre Grenzmatrix und der

letzte Ausdruck gegen Null konvergiert, woraus die Konsistenz von OLS sofort folgt. Man sieht sogar, daß für die Koeffizienten der trend-behafteten Variablen die weit stärkere Aussage $\text{plim } T(\hat{\beta}^{\text{OLS}(b)} - \beta^{(b)}) = 0$ gilt, zuweilen auch Superkonsistenz genannt.

Nun habe aber $\Pi^{(1,2b)}$ keinen vollen Spaltenrang (dabei sei o.B.d.A. die Rangdefizienz = 1). Dann existiert ein $\alpha \neq 0$ mit $\Pi^{(1,2b)}\alpha = 0$. Dabei sei die letzte Komponente von $\alpha = 1$, d.h. $\alpha = (\tilde{\alpha}', 1)'$. Ferner sei \tilde{y} die erste Spalte von $Y^{(1)}$ und $Y^{(1)}$ zerlegt in $Y^{(1)} = (\tilde{y}:\tilde{Y})$, sowie X zerlegt in

$$X = (\tilde{y}:\tilde{Y}:Z^{(1a)}:Z^{(1b)}) \quad (5.41)$$

Man betrachte nun die Regression

$$y^{(1)} = \tilde{X}\delta + u^{(1)} \quad (5.42)$$

mit $\tilde{X} = XP$, $\delta = P^{-1}\beta$ und

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \tilde{\alpha} & 0 & I & 0 \\ -\Pi^{(1,2b)}\alpha & I & 0 & 0 \\ 0 & 0 & 0 & I \end{pmatrix} \quad (5.43)$$

Offenbar ist die Konsistenz von $\hat{\beta}^{\text{OLS}}$ äquivalent zur Konsistenz von $\hat{\delta}^{\text{OLS}}$. Bis auf Spaltenvertauschung unterscheiden sich \tilde{X} und X allein in der ersten Spalte. Deren Komponenten sind weiterhin mit den Störgrößen der Gleichung korreliert, enthalten aber keinen Trend mehr.

In dem zu (5.40) analogen Normalgleichungssystem konvergiert die erste Matrix daher weiterhin gegen einen regulären Grenzwert, aber

$$\text{plim}_{T \rightarrow \infty} \begin{bmatrix} T^{-1} \tilde{X}^{(a)}, u(1) \\ T^{-2} \tilde{X}^{(b)}, u(1) \end{bmatrix} \neq 0 \quad (5.44)$$

(bis auf eine Parametermenge vom Lebesgue-Maß 0). Dies impliziert zunächst

$$\text{plim}_{T \rightarrow \infty} \begin{bmatrix} \hat{\delta}^{\text{OLS}(a)} - \delta^{(a)} \\ T(\hat{\delta}^{\text{OLS}(b)} - \delta^{(b)}) \end{bmatrix} \neq 0, \quad (5.45)$$

damit aber auch $\text{plim} \hat{\delta}^{\text{OLS}(a)} - \delta^{(a)} \neq 0$ (sonst wäre $\text{plim} T(\hat{\delta}^{\text{OLS}(b)} - \delta^{(b)})$ identisch zur Lösung von $\text{plim}(T^{-3} \tilde{X}^{(b)}, X^{(b)}) (T(\hat{\delta}^{\text{OLS}(b)} - \delta^{(b)})) = 0$ und damit ebenfalls $= 0$, ein Widerspruch). q.e.d.

Der nächste Satz beleuchtet die Grenzverteilung von k-Klasse Schätzungen und damit auch von OLS, gegeben OLS ist konsistent.

SATZ (5.4): Für $k = O_p(1)$ und vollen Spaltenrang von $\Pi^{(1,2b)}$ gilt

$$W(\hat{\beta}^{(k)} - \beta) \xrightarrow{D} N(0, \sigma^2_u I). \quad (5.46)$$

Insbesondere haben damit OLS, 2SLS und LIML unter den Bedingungen dieses Kapitels die gleiche Grenzverteilung.

BEWEIS: Der Beweis verläuft nach dem gleichen Muster wie der Beweis von Satz (4.2). Man hat

$$\begin{aligned} W(\hat{\beta}^{(k)} - \beta) &= W(X^{(k)}, X^{(k)})^{-1} X^{(k)}, u(1) \\ &= W(C'Z'ZC)^{-1} C'Z'u(1) \\ &\quad + \{W((X^{(k)}, X^{(k)})^{-1} - (C'Z'ZC)^{-1}) W\}^{-1} X^{(k)}, u(1) \\ &\quad + W(C'Z'ZC)^{-1} W\{W^{-1} X^{(k)}, u(1) - W^{-1} C'Z'u(1)\}. \quad (5.47) \end{aligned}$$

Dabei gilt (vergleiche (5.15) und (5.16))

$$X^{(k)} = ZC + (V^{(1)} - k\hat{V}^{(1)} : 0) = ZC + \tilde{V}^{(k)} \quad (5.48)$$

mit $\tilde{V}^{(k)} = (V^{(1)} - k\hat{V}^{(1)} : 0)$. Falls wie vorausgesetzt $k = O_p(1)$, folgt daraus mit den gleichen Argumenten wie in Lemma (5.1):

$$\text{plim } W(X^{(k)}, X^{(k)})^{-1} W = \lim W(C'Z'ZC)^{-1} = I \quad (5.49)$$

und

$$W^{-1}(\tilde{V}^{(k)}, u^{(1)}) \xrightarrow{P} 0 \quad (5.50)$$

Damit sind die Ausdrücke in geschweiften Klammern asymptotisch vernachlässigbar, und wegen der stochastischen Beschränktheit der übrigen Faktoren verschwinden alle Summanden in (5.47) bis auf den ersten. Dieser ist aber nach Satz (2.3) asymptotisch verteilt wie $N(0, \sigma_u^2 I)$, q.e.d.

Um Satz (5.4) in der herkömmlichen Weise zur Konstruktion von (asymptotischen) Signifikanztests und Konfidenzintervallen für die Regressionskoeffizienten in (5.2) zu verwenden, bleibt wie in Kapitel 4 noch die Konsistenz der OLS-gestützten Schätzung

$$s^2 = (T - g_1 - K_1)^{-1} (y^{(1)} - X\hat{\beta}^{OLS})' (y^{(1)} - X\hat{\beta}^{OLS}) \quad (5.51)$$

für σ_u^2 zu zeigen.

SATZ (5.5): Hat $\Pi^{(1,2b)}$ vollen Spaltenrang, so gilt

$$\text{plim}_{T \rightarrow \infty} s^2 = \sigma_u^2 \quad (5.52)$$

BEWEIS: Man zeigt (5.52) genauso wie Satz (4.3), mit W für D_T , $Z^{(1)}$ für $X^{(a)}$, $Z\Pi^{(1)}$ für $Z^{(b)}$ und $V^{(1)}$ für $V^{(b)}$, q.e.d

Ein weiteres Hindernis vor der praktischen Verwertbarkeit von Satz (5.4) ist die Nichtbeobachtbarkeit des Normalisierungsfaktors W . Sei daher $\tilde{W} = (X'X)^{1/2}$. Im Gegensatz zu W ist \tilde{W} unmittelbar aus den Beobachtungen zu berechnen und kann wegen

$$\tilde{W}^{-1}\tilde{W} \xrightarrow{p} I \quad (5.53)$$

W als Normalisierungsfaktor ersetzen, ohne daß die Grenzverteilung sich ändert. Das entspricht der Ersetzung von D_T durch \tilde{D}_T in Kapitel 4. Damit liefert Satz (5.4) im Verein mit (5.52) und (5.53):

$$\frac{1}{s} \tilde{W} (\hat{\beta}^{\text{OLS}} - \beta) \xrightarrow{D} N(0, I) . \quad (5.54)$$

Das bedeutet, daß auch in Simultanen Gleichungen, sofern nur die Zahl der aus der Gleichung ausgeschlossenen trendbehafteten exogenen Variablen die Zahl der rechtsseitigen endogenen Variablen übersteigt, bei der Konstruktion von Konfidenzintervallen und Signifikanztests für die Regressionsparameter verfahren werden kann wie im Standardregressionsmodell. Zugleich folgt aus Satz (5.4), daß kein anderes k -Klasse Schätzverfahren asymptotisch effizienter ist.

c) Simulation

Die asymptotischen Resultate aus Abschnitt b) liefern keine Anhaltspunkte zu den Stichprobengrößen, ab denen diese Ergebnisse für die ökonometrische Praxis bedeutsam sind. Diese Frage wird hier analog zu Kapitel 3 und 4 mit Hilfe einer Serie von Monte Carlo Experimenten untersucht. Alle Simulationen wurden wiederum mit eigenen Programmen auf dem Sperry Univac 1100 Rechner des Instituts für Höhere Studien durchgeführt, mit Hilfsroutinen aus der NAG-Bibliothek (NUMERICAL ALGORITHMS GROUP, 1981) und der Sperry Univac Math-Pack Bibliothek (SPERRY UNIVAC, 1973) für Zufallszahlenerzeugung und Matrizenoperationen.

Grundlage aller Experimente ist das Modell (5.1), mit $G=2$, $K=5$ und

$$y_{t1} + \beta_{21}y_{t2} + \gamma_{11}z_{t1} + \gamma_{21}z_{t2} + \gamma_{51} = u_{t1} \quad (5.55)$$

$$y_{t1} + \beta_{22}y_{t2} + \gamma_{32}z_{t3} + \gamma_{42}z_{t4} + \gamma_{52} = u_{t2} \quad (5.56)$$

Dabei ist $B = (\beta_{ij})$ und $\Gamma = (\gamma_{ij})$. Die Standardisierung $\beta_{11} = \beta_{12} = 1$ folgt SUMMERS (1965), ebenso wie der übrige Versuchsaufbau. Beide Gleichungen sind wegen zwei in obiger Schreibweise schon berücksichtigter Nullrestriktionen überidentifiziert vom Grad 1.

Das durch (5.55) und (5.56) definierte Modell erlaubt verschiedene ökonomische Interpretationen, etwa als Marktmodell mit einer Angebots- und einer Nachfragegleichung, aus dem die Gleichgewichtsbedingung heraussubstituiert ist, oder als einfaches Keynesianisches

Makromodell, und liegt einer Vielzahl von Monte Carlo-Studien über Schätzverfahren in Simultanen Gleichungen zugrunde (u.a. LADD, 1956; NEISWANGER und YANCEY, 1959; SUMMERS, 1965; SEIBT, 1967; HANSEN, 1969; MIKHAIL, 1975; ATKINSON, 1978). Das gestattet einen unmittelbaren Vergleich der folgenden Ergebnisse mit der einschlägigen Literatur. Ein weiterer Grund für die Wahl dieses einfachen Modells ist die hier im Vordergrund stehende Untersuchung von Trend und die dadurch nötigen großen Stichprobenumfänge. Größere Modelle wären hier sowohl hinsichtlich Speicherbedarf als auch Rechenzeit an EDV-technische Grenzen gestoßen.

Alle Zufallsexperimente beschränken sich auf OLS und 2SLS als den beiden praktisch bedeutsamsten k -Klasse Schätzungen. Dabei wird allein (5.55) explizit geschätzt.

Als Gütekriterium dient wie bisher vor allem die Mittlere Quadratische Abweichung der geschätzten von den wahren Strukturparametern. Daneben liefert das Simulationsprogramm auch empirische Verzerrungen sowie detailliert Histogramme der Schätzwertverteilungen und damit indirekt auch die Möglichkeit zur Ableitung weiterer und insbesondere momentenfreier Qualitätskriterien.

SAWA (1969) beweist im vorliegenden Modellzusammenhang die Existenz des Erwartungswerts und die Nichtexistenz höherer Momente von 2SLS. Allgemein existieren die Momente von 2SLS bis zur Ordnung $K - K_1 - 1$. Der empirische MSE von 2SLS ist daher nicht als Schätzung eines theoretischen Gegenstücks interpretierbar, sondern beschreibt wie in den vorangegangenen Kapiteln nur eine gestutzte Verteilung mit Momenten beliebiger Ordnung. Für OLS hat man diese Interpretationsprobleme nicht -alle Momente existieren bis zur Ordnung $T-1$.

Die Effizienz der Monte Carlo Simulationen wird durch eine von MIKHAIL (1972) vorgeschlagene Kontrollvariable erhöht (siehe dazu auch HENDRY und HARRISON, 1974). Dazu sei die zu schätzende Gleichung allgemein wie in (5.2),

$$\bar{X} = (Z\Pi^{(1)}; Z^{(1)}) \quad (5.57)$$

der deterministische Teil der Regressormatrix X, und

$$\beta^* = \beta + (\bar{X}'Z(Z'Z)^{-1}Z'\bar{X})^{-1}\bar{X}'Z(Z'Z)^{-1}Z'u^{(1)} \quad (5.58)$$

(siehe Mikhail, 1972, S. 622, Gl. 4.4). β^* ist sowohl mit $\hat{\beta}^{OLS}$ als auch $\hat{\beta}^{2SLS}$ positiv korreliert, besitzt aber einen unmittelbar angebbaren Erwartungswert ($= \beta$) und eine ebenso unmittelbar aus (5.58) folgende Kovarianzmatrix

$$\sigma_u^2 (\bar{X}'\bar{X})^{-1} \quad (5.59)$$

Wegen $E(\beta^*) = \beta$ ist (5.59) zugleich die Mittlere Quadratische Abweichung von β^* . Mit $bias(\beta^*)(e)$ und $MSE(\beta^*)(e)$ als den empirischen Gegenstücken hat man dann

$$bias(\beta^*)(e) = \beta \quad (5.60)$$

als Kontrollvariable bei der Simulierung der Verzerrung und

$$MSE(\beta^*)(e) = \sigma_u^2 (\bar{X}'\bar{X})^{-1} \quad (5.61)$$

als Kontrollvariable bei der Simulierung der Mittleren Quadratischen Abweichung, sowohl für OLS als auch 2SLS.

Die Strukturparameter in (5.55) und (5.56) wurden wie in Summers (1965) gewählt, d.h.

$$B = \begin{bmatrix} 1 & 1 \\ \beta_{21} & .4 \end{bmatrix}; \quad \Gamma = \begin{bmatrix} .8 & 0 \\ .7 & 0 \\ 0 & .6 \\ 0 & -.4 \\ -149.5 & -149.5 \end{bmatrix}; \quad \Sigma = \sigma^2 \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}.$$

Ähnlich verfährt auch ATKINSON (1978). Ebenfalls in Anlehnung an Summers variiert β_{21} von $-.1$ über $-.7$ bis -1.3 . σ^2 erhält alternativ die Werte 10 und 100, und die u_{ti} ($i=1,2$) sind wie bei Summers normalverteilt.

Leider erwies sich die Übernahme der exogenen Zeitreihen trotz einer Beschreibung in Summers (1965, S.9) als unmöglich. Wie auch immer die Auslegung der Konstruktionsvorschrift, die resultierenden Reihen hatten nie den Verlauf wie bei Summers (S.8, Schaubild 1 und 2). Stattdessen wurden die z_{ti} ($i=1, \dots, 5; t=1, \dots, T$) durch die folgende Vorschrift erzeugt:

$$\begin{aligned} z_{t1} &= 10 \cdot \sin(t) , \\ z_{t2} &= t , \\ z_{t3} &= t + t \cdot \sin(t) , \\ z_{t4} &= t^2 \quad (z_{t4}^* = t + t \cdot \sin(0.5 \cdot t)) , \\ z_{t5} &= 1 . \end{aligned} \tag{5.62}$$

Einer ersten Versuchsreihe lag die durch $(z_{.5}:z_{.1}:z_{.2}:z_{.3}:z_{.4})$ definierte Z-Matrix zugrunde. Hier hat man

$$\lim_{T \rightarrow \infty} D_T^{-1} Z' Z D_T^{-1} = \begin{bmatrix} 1 & & & & & \\ 0 & 1 & & & & \\ 1/2 & (2/3)^{1/2} & & & & \\ 0 & (15/16)^{1/2} & (10/16)^{1/2} & & & \\ 0 & (3/4)^{1/2} & (1/2)^{1/2} & (5/9)^{1/2} & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix},$$

eine reguläre Matrix (siehe GRENANDER und ROSENBLATT, 1957, S. 245). Die Bedingung (5.20) ist damit erfüllt. Ferner enthalten beide aus der Gleichung ausgeschlossenen exogenen Variablen einen Trend, so daß OLS wegen Satz (5.2) konsistent und wegen Satz (5.4) asymptotisch normalverteilt ist (der volle Spaltenrang von $\Pi^{(1,2b)}$ ist leicht nachzuprüfen).

Die im weiteren unterstellten Stichprobenumfänge sind $T=20, 50, 100$ und 200 . $T=400$ entfällt im vorliegenden Zusammenhang wegen zu großen Rechenzeitaufwands. Tabelle (5.1) gibt $D_T^{-1} Z' Z D_T^{-1}$ für diese T -Werte wieder. Wie man daraus sieht, wird die obige Grenzmatrix schon bei kleinen Stichproben gut approximiert. Die Korrelationsstruktur entspricht dabei durchaus empirisch beobachtbaren Variablenverläufen. Insbesondere sind die zweite und vierte Spalte von Z hoch positiv korreliert.

In einer zweiten Versuchsreihe tritt z_{t4}^* an die Stelle von z_{t4} , d.h. das maximale Diagonalelement von $Z'Z$ ist nur noch $O(T^3)$ und nicht wie vorher $O(T^5)$. Das Analogon zur obigen Grenzmatrix und zu Tabelle (5.1) erhält man durch Übertragung der Korrelationen aus Tabelle (5.2). Auch hier ist $\lim_{T \rightarrow \infty} D_T^{-1} Z' Z D_T^{-1}$ regulär, wie man leicht nachrechnet.

Eine weitere Versuchsreihe schließlich kontrastiert die bei Trend erzielten Simulationsergebnisse mit Schätzungen ohne Trend.

TABELLE (5.1): Korrelation der Regressoren
für verschiedene T

	z_{t5}	z_{t1}	z_{t2}	z_{t3}	z_{t4}
	1				
T = 20	.07	1			
	.88	.01	1		
	.71	.51	.82	1	
	.75	.03	.97	.81	1
	1				
T = 50	.00	1			
	.87	-.05	1		
	.71	.48	.80	1	
	.75	.03	.97	.81	1
	1				
T = 100	.00	1			
	.87	-.03	1		
	.71	.49	.81	1	
	.75	-.03	.97	.77	1
	1				
T = 200	.00	1			
	.87	-.01	1		
	.71	.50	.81	1	
	.75	-.01	.97	.78	1
	1				
T = ∞	.00	1			
	.87	.00	1		
	.71	.50	.82	1	
	.75	.00	.97	.79	1

TABELLE (5.2): Korrelation von z_{t4}^* mit
 z_{ti} ($i \neq 4$)

T	z_{t5}	z_{t1}	z_{t2}	z_{t3}
20	.75	-.05	.88	.66
50	.72	-.04	.79	.63
100	.71	-.02	.80	.65
200	.71	-.01	.81	.66
∞	.71	.00	.82	.67

Dazu wurden die ersten 20 Z-Werte (mit z_{t4}^* für z_{t4}) für $T > 20$ sukzessive wiederholt.

Die in den folgenden Experimenten variierten Größen sind damit Z (3 Ausprägungen), T (4 Ausprägungen), β_{21} (3 Ausprägungen) und σ^2 (2 Ausprägungen). Jede Kombination dieser Parameter definiert ein Experiment, insgesamt $4 \times 3 \times 3 \times 2 = 72$. Pro Experiment wurden mit jeweils neuen Zufallszahlen $N=100$ Schätzungen simuliert, zur genaueren Inspektion der empirischen Verteilung zuweilen aber auch 10 000.

Tabelle (5.3) gibt in komprimierter Form die Ergebnisse der ersten Versuchsreihe wieder. Es werden allein die Quotienten der empirischen Mittleren Quadratischen Abweichungen explizit aufgeführt.

Die Tabelle zeigt schon für $T=100$ keinen großen Qualitätsunterschied mehr zwischen OLS und 2SLS. Für kleines T ist die Rangfolge der Verfahren für gegebene β_{21} und σ^2 nicht über alle Strukturparameter identisch, eine Standardbeobachtung auch anderer Monte Carlo Simulationen im vorliegenden Modellzusammenhang. Insbesondere ist OLS auch für kleine T mitunter besser (nach dem MSE-Kriterium) als 2SLS, in Übereinstimmung mit den Monte Carlo Ergebnissen von SUMMERS (1965, S.29), QUANDT (1965, S. 101), MIKHAIL (1975, S. 96) oder PARK (1982), oder den analytischen Berechnungen von SAWA (1969) und RICHARDSON und WU (1971). Mit wachsendem T tritt dann aber eine zunehmende Angleichung der Verfahren ein. Ebenfalls in Übereinstimmung mit bekannten Resultaten ist der im Vergleich zu 2SLS größte Qualitätsabfall von OLS bei der Schätzung des Koeffizienten der erklärenden endogenen Variablen der Gleichung.

TABELLE (5.3): Relative Mittlere Quadratische Abweichungen
 $(z_{t4} = t^2)$

T	σ^2							
	10				100			
	γ_{51}	γ_{11}	γ_{21}	β_{21}	γ_{51}	γ_{11}	γ_{21}	β_{21}
	a) $\beta_{21} = -.1$							
20	1.36	.88	1.68	1.72	1.50	.43	2.28	2.29
50	1.16	.99	1.26	1.29	1.99	.93	2.65	2.77
100	1.01	1.00	1.03	1.03	1.09	.98	1.21	1.22
200	1.01	1.00	1.02	1.02	1.04	1.00	1.07	1.08
	b) $\beta_{21} = -.7$							
20	1.16	.80	1.47	1.53	1.34	.38	1.72	1.74
50	1.10	.99	1.19	1.21	1.78	.95	2.36	2.44
100	1.09	1.00	1.08	1.06	1.37	.99	1.39	1.34
200	1.00	1.00	1.00	1.00	1.01	1.00	1.04	1.03
	c) $\beta_{21} = -1.3$							
20	1.56	.93	2.02	2.13	1.79	.52	2.31	2.32
50	1.00	1.00	1.05	1.07	1.51	.96	2.00	2.13
100	1.03	1.00	1.03	1.03	1.19	.99	1.24	1.24
200	1.00	1.00	1.01	1.01	1.02	1.00	1.04	1.04

Tabelle (5.4) schlüsselt die in Tabelle (5.3) sehr komprimierte Information beispielhaft für $T=20$, $\sigma^2=100$ und $\beta_{21} = -1.3$ näher auf. Verzerrung und MSE erscheinen getrennt, sowohl bei direkter Simulation (d) als auch bei Verwendung der oben definierten Kontrollvariablen. In der Zeile (se) sind die jeweiligen Standardfehler der

Monte Carlo Resultate aufgeführt (zu deren Berechnung siehe MIKHAIL, 1972, S. 623). Das erlaubt die Bestimmung des bei der Verwendung von Kontrollvariablen erzielten Effizienzgewinns (bzw. -verlusts). Der Quotient der Standardabweichungen von direkten und mit Kontrollvariablen korregierten Simulationsergebnissen ist in der Zeile (eff) aufgeführt.

Man sieht zunächst, daß die hier gewählte Kontrollvariable nicht in jedem Fall einen Effizienzgewinn bedeutet. Nur unter (4.55) kann man mit Kontrollvariablen eine Effizienzsteigerung erwarten, und diese Bedingung ist offenbar bei den hier zugrundegelegten Parametern für OLS verletzt. Für 2SLS wird die Simulationseffizienz für die Verzerrung etwa verdoppelt, d.h. zur Erzielung der gleichen Simulationsgenauigkeit wären bei direkter Simulation doppelt so viele Wiederholungen nötig. Wie aber weiter unter Tabelle (5.5) zeigt, wird dieser Effizienzgewinn für andere Parameterkonstellationen noch weit übertroffen und erreicht auch für OLS und auch für die Monte Carlo Schätzung der Mittleren Quadratischen Abweichung Werte von mehr als 20. Die Quotienten der Mittleren Quadratischen Abweichungen in Tabelle (5.3) und weiter unten basieren auf den jeweils genaueren Simulationsergebnissen.

Tabelle (5.4) dokumentiert auch die im Vergleich zu 2SLS weitaus größere Verzerrung von OLS. Diese wird aber mitunter (etwa bei γ_{11}) durch die kleinere Varianz mehr als ausgeglichen. Die empirischen Varianzen sind in Tabelle (5.4) nicht explizit aufgeführt, können aber indirekt über

$$\text{MSE} = (\text{bias})^2 + \text{var} \quad (5.63)$$

TABELLE (5.4): Verzerrung und Mittlere Quadratische Abweichung bei direkter und Kontrollvariablen-Simulation (T=20)

		γ_{51} (-149.5)	γ_{11} (.8)	γ_{21} (.7)	β_{21} (-1.3)
a) OLS					
bias	(d)	-9.0014	.1789	2.5865	-.5274
	(se)	.3138	.0183	.0596	.0115
	(cv)	-8.4476	.1586	2.3803	-.4765
	(se)	.5894	.0193	.1521	.0303
	(eff)	1.8785	1.0569	2.5505	2.3258
MSE	(d)	90.8713	.0655	7.0457	.2914
	(se)	5.8065	.0069	.3216	.0125
	(cv)	96.7194	.0728	7.3412	.2914
	(se)	8.0850	.0114	.4162	.0159
	(eff)	1.3924	1.6491	1.2940	1.2666
b) 2SLS					
bias	(d)	-.8203	.0234	.2767	-.0644
	(se)	.6659	.0345	.1638	.0335
	(cv)	-.2665	.0013	.0705	-.0135
	(se)	.3531	.0138	.0957	.0197
	(eff)	.5302	.3986	.5841	.5893
MSE	(d)	45.0153	.1196	2.7607	.1163
	(se)	9.5160	.0195	.6532	.0292
	(cv)	50.8632	.1270	3.0561	.0292
	(se)	8.5876	.0169	.5777	.0258
	(eff)	.9024	.8681	.8843	.8839

ermittelt werden.

Die Richtung der Verzerrung ist für OLS und 2SLS und für alle Regressionskoeffizienten gleich dem Vorzeichen der Korrelation von endogenem Regressor und Störgröße, in Bestätigung eines bekannten analytischen Resultats (siehe etwa MARIANO, 1982, S. 523). Im vorliegenden Zusammenhang gilt

$$\text{cov}(-y_{t2}, u_{t1}) = \sigma^2 \frac{0.5 - \beta_{21}}{0.4 - \beta_{21}} \quad . \quad (5.64)$$

Dieser Ausdruck ist für $\beta_{21} = -1.3$ positiv, d.h. für alle k-Klasse Schätzungen mit festem k ($0 \leq k \leq 1$) liegt der Erwartungswert bei negativen Koeffizienten rechts und bei positiven Koeffizienten links vom wahren Parameterwert (die empirischen Verzerrungen in Tabelle (5.4) wurden aus der Regressionschreibweise von (5.55) gewonnen, mit allen erklärenden Variablen rechts des Gleichheitszeichens, und sind daher im Sinn der Schreibweise (5.55) vorzeichenverkehrt).

Tabelle (5.5) gibt die analogen Ergebnisse für T=50 wieder. Sowohl Verzerrung als auch MSE sind für beide Verfahren nun drastisch geringer, genauso wie der Qualitätsunterschied. Ferner fällt der nun auch bei OLS durch Kontrollvariablen erzielbare Effizienzgewinn bei der Simulation auf, eine Folge der zunehmenden Korrelation von Schätzfunktion und Kontrollvariablen. Nicht verwundern darf dabei der im Vergleich zur Verzerrung geringere Effizienzgewinn bei der Simulation der Mittleren Quadratischen Abweichung: die Korrelation von (5.61) mit den direkt simulierten Mittleren Quadratischen Abweichungen von OLS und 2SLS ist kleiner als die entsprechenden Werte bei der Verzerrung (Mikhail, 1972, S. 622).

TABELLE (5.5): Verzerrung und Mittlere Quadratische Abweichung bei direkter und Kontrollvariablen-Simulation (T=50)

		γ_{51} (-149.5)	γ_{11} (.8)	γ_{21} (.7)	β_{21} (-1.3)
a) OLS					
bias	(d)	-3.7834	-.0199	.4421	-.0371
	(se)	.4108	.0169	.0368	.0030
	(cv)	-4.0319	-.0169	.4695	-.0390
	(se)	.1112	.0009	.0127	.0011
	(eff)	.2707	.0503	.3454	.3548
MSE	(d)	31.1867	.0290	.3313	.0023
	(se)	3.6390	.0048	.0362	.0002
	(cv)	29.2517	.0391	.3157	.0022
	(se)	3.4669	.0007	.0358	.0002
	(eff)	.9527	.1390	.9909	.9607
b) 2SLS					
bias	(d)	.2508	-.0038	-.0271	.0019
	(se)	.4605	.0175	.0416	.0033
	(cv)	.0023	-.0008	.0003	.0000
	(se)	.0267	.0006	.0027	.0002
	(eff)	.0580	.0331	.0660	.0668
MSE	(d)	21.2717	.0307	.1735	.0011
	(se)	3.4568	.0050	.0250	.0001
	(cv)	19.3767	.0407	.1579	.0010
	(se)	.5043	.0003	.0038	.0000
	(eff)	.1459	.0684	.1503	.1526

Zur weiteren Dokumentation des unterschiedlichen Gewichts von Verzerrung und Varianz als Komponenten der Mittleren Quadratischen Abweichung zeigt Schaubild (5.1) Histogramme der OLS- und 2SLS-Schätzungen von β_{21} für die Parameterkonstellation aus Tabelle (5.4). Dazu wurde das Experiment mit 10 000 Durchgängen wiederholt. Das Resultat zeigt deutlich die dichte Konzentration von OLS um den Erwartungswert und die bei 2SLS vergleichsweise größere Varianz, bei gleichzeitig geringerer Verzerrung.

SCHAUBILD (5.1): Histogramme von je 10 000 Realisationen von OLS und 2SLS ($-\beta_{21}=1.3$; $T=20$; $\sigma^2=100$)

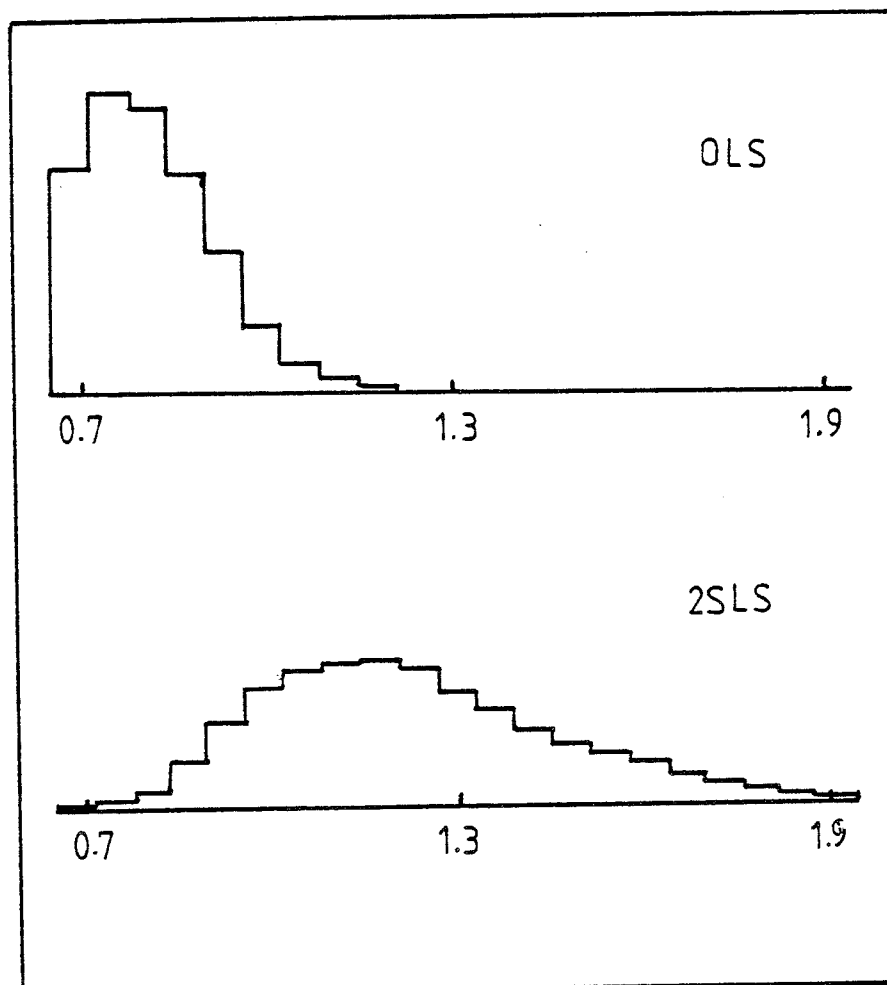


Tabelle (5.6) reproduziert die zu Tabelle (5.3) analogen Simulationsergebnisse für den Fall der Z-Matrix ohne Trend. Wie nicht anders zu erwarten, sinkt die relative Qualität von OLS wegen nicht verschwindender Verzerrung mit wachsendem T drastisch ab. Für T=20 ist OLS aber noch häufig überlegen, und diese

TABELLE (5.6): Relative Mittlere Quadratische Abweichungen
(kein Trend)

T	σ^2							
	10				100			
	γ_{51}	γ_{11}	γ_{21}	β_{21}	γ_{51}	γ_{11}	γ_{21}	β_{21}
	a) $\beta_{21} = -.1$							
20	.36	.39	.98	2.18	.16	.29	.28	.55
50	.40	1.08	2.54	6.87	.14	.41	.44	1.25
100	.60	1.01	4.52	12.88	.19	.42	.53	1.66
200	.84	1.43	10.24	32.22	.35	.43	2.30	6.97
	b) $\beta_{21} = -.7$							
20	.35	.38	.91	1.58	.20	.18	.42	.66
50	.46	.86	2.58	7.14	.24	.29	.51	1.00
100	.54	.99	5.05	17.74	.26	.33	1.03	3.34
200	.74	1.82	10.48	30.13	.25	.66	1.64	4.74
	c) $\beta_{21} = -1.3$							
20	.38	.48	1.22	2.89	.28	.47	.68	1.20
50	.55	.82	1.73	5.29	.18	.31	.38	.86
100	.61	.84	5.37	15.38	.38	.28	.49	.71
200	.66	1.60	10.23	31.58	.32	.51	2.39	7.38

Relative Überlegenheit nimmt ceteris paribus mit wachsender Störgrößenvarianz zu. Ferner ist die Verzerrung der OLS-Schätzung für die Konstante γ_{51} so gering, daß hier die Überlegenheit sogar bis $T=200$ bestehen bleibt.

Eine letzte Versuchsreihe gilt der durch Einsetzen von z_{t4}^* an Stelle von z_{t4} gewonnenen Z-Matrix. Durch den quadratischen Trend in z_{t4} sind sowohl $\|z_{.4}\|$ als auch $\|y_{.2}\|$ von der Ordnung T^5 , ein möglicherweise unrealistisch schnelles Wachstum. Dadurch werden die Auswirkungen eines weniger drastischen Trends unter Umständen überdeckt. Mit z_{t4}^* an der Stelle von z_{t4} ist $Z'Z$ nunmehr von der Ordnung T^3 .

Tabelle (5.7) gibt die zu Tabelle (5.3) und Tabelle (5.6) analogen Simulationsergebnisse wieder. Dabei sind die Zeilen $T=20$ wegen identischer Z-Matrizen aus Tabelle (5.6) übernommen. Wie aber die Ergebnisse für $T=50$ und $T=100$ zeigen, nimmt die relative Effizienz von OLS nun mit wachsendem T zunächst ab. Das entspricht den Resultaten in Tabelle (4.2) für das Fehler-in-den-Variablen Modell. Der Grund ist wie in Schaubild (4.2) das vergleichsweise langsame Verschwinden der Mittleren Quadratischen Abweichung von OLS. Besonders deutlich wirkt sich das für $\sigma^2=100$ aus. Bei der Schätzung von β_{21} in Tabelle (5.7a) etwa sinkt $MSE(OLS)$ von .0559 bei $T=20$ auf .0015 bei $T=200$. Auf der anderen Seite sind die entsprechenden Werte für 2SLS .1022 ($T=20$) und .0001 ($T=200$), mit dem Resultat einer Steigerung des Quotienten der Mittleren Quadratischen Abweichungen von .55 auf 20.75. Der Punkt T_2^* der minimalen relativen Effizienz von OLS aus Schaubild (4.2) scheint dabei mit wachsendem σ^2 und abnehmendem Trend nach rechts zu rücken,

TABELLE (5.7): Relative Mittlere Quadratische Abweichungen

$$(z_{t4} = t + t \sin(.5 t))$$

T	σ^2							
	10				100			
	γ_{51}	γ_{11}	γ_{21}	β_{21}	γ_{51}	γ_{11}	γ_{21}	β_{21}
	a) $\beta_{21} = -.1$							
20	.36	.39	.98	2.18	.16	.29	.28	.55
50	.82	2.00	2.31	6.57	.33	1.79	2.00	5.62
100	.92	1.99	2.08	4.91	.68	6.09	4.96	15.89
200	.98	2.54	1.92	3.95	.82	10.22	6.89	20.75
	b) $\beta_{21} = -.7$							
20	.35	.38	.91	1.58	.19	.18	.42	.66
50	.78	1.86	1.94	5.86	.33	1.47	1.49	5.33
100	.85	2.43	2.73	5.55	.51	6.82	6.57	16.38
200	.98	1.96	1.77	3.15	.80	8.84	6.79	18.85
	c) $\beta_{21} = -1.3$							
20	.38	.48	1.23	2.89	.28	.47	.68	1.20
50	.85	1.97	1.86	6.52	.39	1.44	1.35	5.17
100	.92	2.41	2.24	5.34	.54	6.70	5.65	17.65
200	.98	2.15	1.78	3.34	.86	9.18	6.43	18.98

bei gleichzeitiger Abnahme der relativen Effizienz selbst. Wie schon in Kapitel 3 und Kapitel 4 bleibt daher festzustellen, daß die asymptotische Äquivalenz von OLS zu den jeweils konkurrierenden Verfahren unter Umständen erst bei sehr großen Stichprobenumfängen wirksam wird.

Eine abschließende Serie von 3 Experimenten hat die Überprüfung der approximativen Normalverteilung des normierten Schätzfehlers von OLS zum Ziel. Wegen (5.54) hat man z.B.

$$(s^2 a_{y_2})^{-1/2} (\hat{\beta}_{21}^{\text{OLS}} - \beta_{21}) \xrightarrow{D} N(0,1) . \quad (5.65)$$

Dabei ist a_{y_2} das zu y_{t2} gehörende Element der Inversen der Kreuzproduktmatrix der Regressoren aus (5.55). Im weiteren wird die Approximationsqualität von (5.65) mit den Parameterwerten aus Tabelle (5.3c) nachgerechnet. Dabei wird $\sigma^2=10$ gesetzt, und die Stichprobengröße $T=200$ entfällt wegen exzessiven Rechenzeitbedarfs. Für $T=20, 50$ und 100 werden die entsprechenden Experimente aus Tabelle (5.3c) mit jeweils 10 000 Durchgängen wiederholt.

Schaubild (5.2) gibt wie Schaubild (4.4) im Fehler-in-den-Variablen Modell die Histogramme der resultierenden normierten Schätzfehler an, jeweils wiederum mit der Dichtefunktion der Standardnormalverteilung zum besseren Vergleich der Approximationsqualität. Diese erscheint in der Tat sehr unbefriedigend, nimmt aber mit wachsendem T der Theorie entsprechend zu.

Mit den gleichen Simulationsergebnissen läßt sich auch das wahre Signifikanzniveau für den Standardtest der Hypothese $H_0: \beta_{21} = -1.3$ nachprüfen. Tabelle (5.8) vergleicht analog zu Tabelle (4.5) einige nominelle mit den zugehörigen tatsächlichen Signifikanzniveaus. Die ungewohnten nominellen Signifikanzniveaus sind dabei wiederum eine Folge der Intervallgrenzen der Histogramme in Schaubild (5.2).

SCHAUBILD (5.2): Histogramme von je 10 000 Realisationen des normierten Schätzfehlers von OLS ($\beta_{21} = -1.3$, $\sigma^2 = 10$)

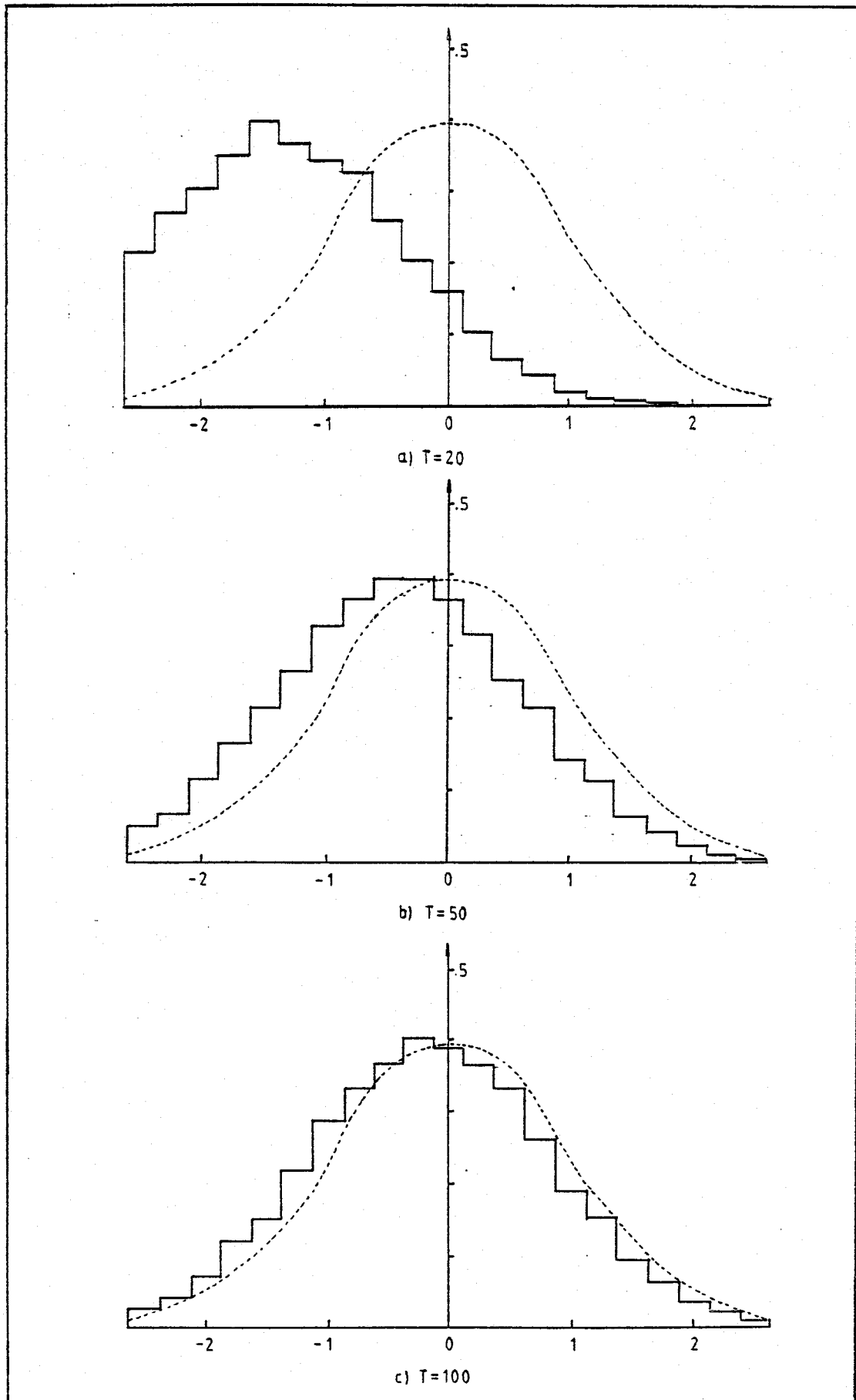


TABELLE (5.8): Vergleich von nominellem und tatsächlichem Signifikanzniveau beim Test von $H_0: \beta_{21} = -1.3$ ($z_{t4} = t^2$)

	H_1					
	$\beta_{21} \geq -1.3$		$\beta_{21} \leq -1.3$		$\beta_{21} \neq -1.3$	
$\alpha(\text{nominell}):$	5.2%	.88%	5.2%	.88%	10.4%	1.7%
T= 20	41.5%	18.5%	.2%	.03%	41.7%	18.5%
$\alpha(\text{real}):$ T= 50	12.7%	2.9%	2.2%	.3%	14.9%	3.2%
T=100	7.1%	1.4%	3.7%	.6%	10.8%	2.0%

Die Ergebnisse bestätigen den visuellen Eindruck aus Schaubild (5.2): Für $T=20$ ist (5.65) als Grundlage eines Signifikanztests völlig ungeeignet, und auch bei wachsendem T wird das nominelle Signifikanzniveau nur zögernd approximiert. Das bestätigt für OLS die 2SLS-orientierten analytischen Ergebnisse von RICHARDSON und ROHR (1971), sowie die Monte Carlo Resultate von BASMAN et al. (1974a,b) und MADDALA (1974). Für andere Parameterkonstellationen ($\sigma^2=100$, linearer statt quadratischer Trend) ergab sich sogar eine noch schlechtere Normalverteilungsapproximation, soweit aus nur 100 Durchgängen beurteilbar. Diese Ergebnisse können daher als generelle Warnung verstanden werden, in Simultanen Gleichungen einer asymptotischen Normalverteilungsapproximation nur bei sehr großen Stichproben zu vertrauen.

d) Diskussion

BASMAN (1963) und RICHARDSON (1968) geben die exakte Dichte und Verzerrung der 2SLS-Schätzung für β_{21} im Modell (5.55)-(5.56) an, gegeben normalverteilte u_{ti} ($i=1,2$). RICHARDSON und WU (1971) und unabhängig davon SAWA (1969) verallgemeinern diese Resultate auf MSE und OLS. Verallgemeinerungen auf Gleichungen mit mehr als einer rechtsseitigen endogenen Variablen liefern ULLAH und NAGAR (1974) sowie HILLIER, KINAL und SRIVASTAVA (1984). Nützliche Übersichten dazu findet man bei BASMAN (1974), PHILLIPS (1980), MARIANO (1982) oder TAYLOR (1983).

Wegen des rechentechnischen Aufwands bei der Anwendung exakter Ergebnisse, der Restriktionen sowohl bezüglich Störgrößenverteilungen als auch zugelassener Schätzverfahren, der Beschränkung auf korrekt spezifizierte Modelle, sowie auch des Mangels an Ergebnissen für Modelle mit verzögerten endogenen Variablen bleiben aber Monte Carlo Studien auch weiterhin eine verbreitete Untersuchungsmethode im Kontext Simultaner Gleichungssysteme. Jüngere Beispiele findet man etwa bei ATKINSON (1978), PARK (1982), FREIMANN (1982), FREIMANN und GRUBER (1983), oder KRÄMER (1983). Eine Zusammenstellung der wichtigsten bis dato bekannten Arbeiten liefert JOHNSTON (1972, Kap. 11), und tabellarische Übersichten findet man bei SMITH (1971) und KRÄMER (1980a, S. 51).

Eine zusätzliche Existenzberechtigung für Monte Carlo-Untersuchungen im vorliegenden Modellzusammenhang ist der weitgehende Mangel an exakten Aussagen zu Verteilungen von Teststatistiken. RICHARDSON und ROHR (1971) geben die exakte Verteilung des 2SLS-

Analogons von (5.65) im Kontext des einfachen statischen Modells mit zwei endogenen Variablen an, aber vergleichbare Ergebnisse für andere Verfahren und allgemeinere Modelle fehlen bisher (nicht gerechnet die Approximationen in SARGAN, 1975, 1976).

KRÄMER (1981) untersucht OLS-Schätzungen bei linearem Trend und beschränkten Abweichungen vom Trend. Diese Situation ist hier bei mehr als einer trendbehafteten Variablen durch die Annahme (5.20) ausgeschlossen, und führt zu vergleichsweise restriktiveren Bedingungen für die Konsistenz und asymptotische Normalverteilung von OLS.

Satz (5.4) zeigt allein die durch Trend möglicherweise bewirkte asymptotische Äquivalenz von k-Klasse Schätzungen, d.h. Schätzverfahren bei beschränkter Information. Es bleibt die Frage, ob System-Schätzverfahren, die auch Informationen aus anderen Gleichungen verwerten, nicht auch weiterhin asymptotisch überlegen sind. Wie das folgende einfache Beispiel zeigt, ist dies in der Tat der Fall:

Es sei

$$\begin{aligned} y_t^{(1)} &= x_t^{(1)} \beta^{(1)} + u_t^{(1)} \\ y_t^{(2)} &= x_t^{(2)} \beta^{(2)} + u_t^{(2)}, \quad t=1,2,\dots \end{aligned} \quad (5.66)$$

mit $E(u_t^{(i)}) = 0$, $E(u_t^{(i)2}) = 1$ ($i=1,2$), $E(u_t^{(1)} u_t^{(2)}) = \sigma_{12} \neq 0$, und skalaren, exogenen $x_t^{(i)}$. Damit definiert (5.66) ein "seemingly unrelated regression" (SUR)-System. Wegen der Exogenität der $x_t^{(i)}$ ist OLS hier identisch mit 2SLS, und die von ZELLNER (1962) eingeführte SUR-Schätzung entspricht "Three-Stage-Least-Squares" (3SLS). $\hat{\beta}^{(1)OLS}$ und $\hat{\beta}^{(1)3SLS}$ seien die entsprechenden Schätzwerte für

$\beta^{(1)}$. Mit $x^{(i)} := (x_1^{(i)}, \dots, x_T^{(i)})'$ und für seriell unkorrelierte $u_t^{(i)}$ rechnet man leicht nach, daß

$$\text{var}(\hat{\beta}^{(1)\text{OLS}}) = \text{MSE}(\hat{\beta}^{(1)\text{OLS}}) = (x^{(1)'} x^{(1)})^{-1} \quad (5.67)$$

und

$$\begin{aligned} \text{var}(\hat{\beta}^{(1)\text{3SLS}}) &= \text{MSE}(\hat{\beta}^{(1)\text{3SLS}}) \\ &= \frac{(1 - \sigma_{12}^2) x^{(2)'} x^{(2)}}{x^{(1)'} x^{(1)} \cdot x^{(2)'} x^{(2)} - \sigma_{12}^2 (x^{(1)'} x^{(2)})^2}. \end{aligned} \quad (5.68)$$

Für $\sigma_{12} = 0$ ist $\hat{\beta}^{(1)\text{OLS}} = \hat{\beta}^{(1)\text{2SLS}}$ und (5.67) identisch mit (5.68).

Für $\sigma_{12} \neq 0$ und $x^{(2)}$ verschieden von einem Vielfachen von $x^{(1)}$ ist (5.68) aber kleiner als (5.67), und der Quotient geht auch bei Trend in den exogenen Variablen im allgemeinen nicht gegen 1.

Man betrachte etwa den Fall $x_t^{(1)} = t$, $x_t^{(2)} = t^2$. Hier gilt

$$T^3 \text{var}(\hat{\beta}^{(1)\text{OLS}}) \longrightarrow 3, \quad (5.69)$$

$$T^3 \text{var}(\hat{\beta}^{(1)\text{3SLS}}) \longrightarrow 3 \frac{1 - \sigma_{12}^2}{1 - \frac{15}{16} \sigma_{12}^2}, \quad (5.70)$$

d.h. der Effizienzvorsprung von 3SLS verschwindet auch asymptotisch nicht.

6. ZUSAMMENFASSUNG

Die vorliegende Arbeit hat für verschiedene Modellzusammenhänge Bedingungen angegeben, unter denen bei Trend in den exogenen Variablen die ansonsten vorliegende Unterlegenheit von OLS im Vergleich zu jeweiligen Konkurrenzverfahren zumindest asymptotisch verschwindet (im Sinn identischer Grenzverteilungen der normierten Schätzfehler). Zugleich werden in diesen Fällen auch die Unterschiede zwischen anderen Verfahren nivelliert. Auch wenn bei Approximationen höherer Ordnung weiterhin Unterschiede bestehen bleiben (wie etwa in SCHNEEWEISS, 1982, für Fehler in den Variablen diskutiert), kann man Trend doch in dem Ausmaß, wie die Annahmen dieser Arbeit durch die ökonomische Realität gedeckt sind, als den "großen Gleichmacher" ökonomischer Schätzverfahren bezeichnen.

Allerdings sind die Wirkungen von Trend nicht in allen Modellen gleich. Im Standardregressionsmodell mit korrelierten Störgrößen (Kapitel 3) wird OLS durch Trend von einem lediglich konsistenten zu einem relativ zu GLS effizienten Verfahren. Die ausschlaggebende Wirkung von Trend ist hier allein die gegen 1 strebende empirische Autokorrelation beliebiger Ordnung der Regressoren, eine auch ohne Trend nicht ausgeschlossene wie mit Trend nicht notwendig gegebene Situation. Auf der anderen Seite hängen die Resultate in den Kapiteln 4 und 5 entscheidend von der Ordnung der empirischen Momente der Regressoren ab, und wird OLS von einem ansonsten inkonsistenten zu einem konsistenten Verfahren. Die Autokorrelationsstruktur der exogenen Variablen geht darin

an keiner Stelle ein. Die ausschlaggebende Eigenschaft ist hier allein die bei Trend im Vergleich zu den Störgrößen überproportional wachsende empirische Varianz der Regressoren. Insofern liegen den Ergebnissen in Kapitel 3 auf der einen und in den Kapiteln 4 und 5 auf der anderen Seite grundsätzlich verschiedene Zusammenhänge zugrunde.

Verallgemeinerungen der Ergebnisse aus den Kapiteln 4 und 5 bieten sich zunächst an für Modelle mit verzögerten endogenen Variablen unter den Regressoren. Diese sind in der vorliegenden Arbeit vor allem wegen beweistechnischer Schwierigkeiten ausgeschlossen. Auch die Forderung einer konstanten Störgrößenvarianz bei gleichzeitig ungeschränkt wachsenden Regressoren erscheint für viele Resultate unnötig restriktiv. Wie man leicht durch Nachvollziehen der entsprechenden Beweise überprüft, ist etwa für die Konsistenz von OLS bereits die Bedingung hinreichend, daß die empirischen Momente der Regressoren schneller wachsen als diejenigen der Störgrößen.

Schließlich läßt sich die vorliegende Arbeit als weiterer Beitrag zu einer Rehabilitation von OLS auch außerhalb des Standard-Regressionsmodells begreifen. Frühere Ansätze haben etwa die Prognosequalitäten von OLS betont (WAUGH, 1961) oder auf die relative Robustheit von OLS gegenüber verschiedenen Arten von Fehlspezifikation abgestellt, sei es mit Hilfe von Monte Carlo-Simulationen (CRAGG, 1968; KRÄMER, 1980a, 1983c) oder analytischen Betrachtungen (HALE, MARIANO und RAMAGE, 1980; MARIANO und RAMAGE, 1982; RHODES und WESTBROOK, 1984). Zu diesen Überlegungen tritt nun eine auch asymptotisch gültige Rechtfertigung für den Fall von Trend.

Wie die Monte Carlo-Simulationen aber zeigen, erfolgt die asymptotische Angleichung der relativen Qualität von OLS keineswegs monoton. Das gilt im Standardregressionsmodell mit korrelierten Störgrößen (siehe die Tabellen (3.5) und (3.6)), bei Fehlern in den Variablen (siehe die Tabellen (4.2) und (4.3)), und auch in Simultanen Gleichungen (siehe die Tabellen (5.3) und (5.7)). Bevor die asymptotischen Resultate wirksam werden, muß zuweilen ein langes Qualitätstal durchschritten werden, und die nötigen Stichprobenumfänge bleiben oft jenseits praktisch relevanter Größenordnungen. Dies mag abschließend als Warnung gelten, auch bei asymptotischer Äquivalenz von OLS die jeweiligen Konkurrenzverfahren nicht leichtfertig aufzugeben.

ANHANG



a) EMPIRISCHE ANALYSE DER KORRELATION AUSGEWÄHLTER
ÖKONOMISCHER ZEITREIHEN

Dieser Teil des Anhangs diskutiert die Realitätsnähe der Annahmen (2.7) und (2.8) (bzw. (4.14) und (4.15), oder (5.19) und (5.20)) bei trendbehafteten ökonomischen Variablen. Wie schon in den Kapiteln 4 und 5 betont, erzwingt die Koexistenz von Trend und einer regulären Grenzmatrix der normierten Kreuzprodukte (im Sinn von (4.15) und (5.20)) gewisse Restriktionen bezüglich der hier zugelassenen Zeitreihen, wie etwa bei linearem Trend im Zeitablauf wachsende Abweichungen vom Trend. Auf den folgenden Seiten wird an Hand ausgewählter trendbehafteter ökonomischer Zeitreihen aus Österreich die Plausibilität dieser Voraussetzung überprüft.

Natürlich können Annahmen zum Grenzverhalten eines Zeitreihenvektors und daraus abgeleiteter Größen wie normierter Kreuzproduktmatrizen nie aus endlichen Stichproben verifiziert oder widerlegt werden. Man kann aber nachprüfen, ob die aus endlichen Stichproben gewonnenen Größen zulässigen Grenzwerten zustreben oder nicht. Insbesondere wären die in dieser Arbeit getroffenen Annahmen als wenig realitätsnah einzustufen, wenn die normierten Kreuzproduktmatrizen typischer trendbehafteter Regressorenvektoren im Zeitablauf eine Tendenz zur Singularität aufwiesen, wie im Beispiel (4.18).

Die im weiteren untersuchten Zeitreihen sind der Datenbank des Österreichischen Wirtschaftsforschungsinstituts (WIFO) entnommen (siehe WANG, 1978), unter Verwendung des am Institut für Höhere Studien entwickelten ökonometrischen Programmpakets IAS-SYSTEM (siehe PLASSER, 1982, oder SONNBERGER et al., 1983). Aus den mehr als tausend österreichischen und internationalen ökonomischen Zeitreihen der WIFO-Datenbank wurde dabei nach den Kriterien

(i) Länge, (ii) Trend und (iii) Verfügbarkeit ausgewählt. Um möglichst lange Zeitreihen zu erhalten, wurden daher zunächst nur Monatsdaten zugelassen (Wochen- oder Tagesdaten sind in der WIFO-Datenbank nicht vorhanden). Diese werden allerdings erst seit Mitte der 50'er Jahre in größerem Umfang erhoben und sind teilweise noch nicht bis in die jüngste Vergangenheit abgespeichert. Als Kompromiß wurde der Zeitraum Jan. 1954-Dez. 1982 gewählt, entsprechend einem Stichprobenumfang $T=348$. Auf Grund des Kriteriums (ii) blieben alle mit einer natürlichen Obergrenze versehenen Variablen wie Zinssätze oder Arbeitslosenraten ausgeschlossen, und auf Grund des Verfügbarkeitskriteriums blieben eine Anzahl weiterer derzeit wegen Überarbeitung gesperrter Reihen außer Betracht.

Tabelle (A.1), mit der im weiteren verwendeten Kurzschreibweise in der ersten Spalte, gibt eine repräsentative Auswahl der verbleibenden Variablen wieder. Weitere zulässige aber hier nicht berücksichtigte Kandidaten sind sachlich disaggregierte Ex- und Importzeitreihen, zusätzliche Steuerarten (Einkommenssteuer, Umsatzsteuer usw.), sowie Outputreihen alternativer Industriezweige.

Schaubild (A.1) illustriert beispielhaft einige typische Kurvenverläufe von Zeitreihen, die als potentielle exogene Variable in Modellen des Spar- und Konsumverhalten privater Haushalte gelten können. Aus Platzmangel und weil das IAS-SYSTEM (Level 2-18) nur Zeitreihen von maximal 120 Werten zum Plotten zuläßt, sind dabei jeweils nur die ersten 10 Jahre (Jan. 1954-Dez. 1963) wiedergegeben. Zusätzlich sind hier und im weiteren alle monetären Reihen zur Vermeidung exponentiellen Wachstums mit dem Lebenshaltungskostenindex (Apr.1945=100; WIFO-Kennziffer PVLH4S) deflationiert).

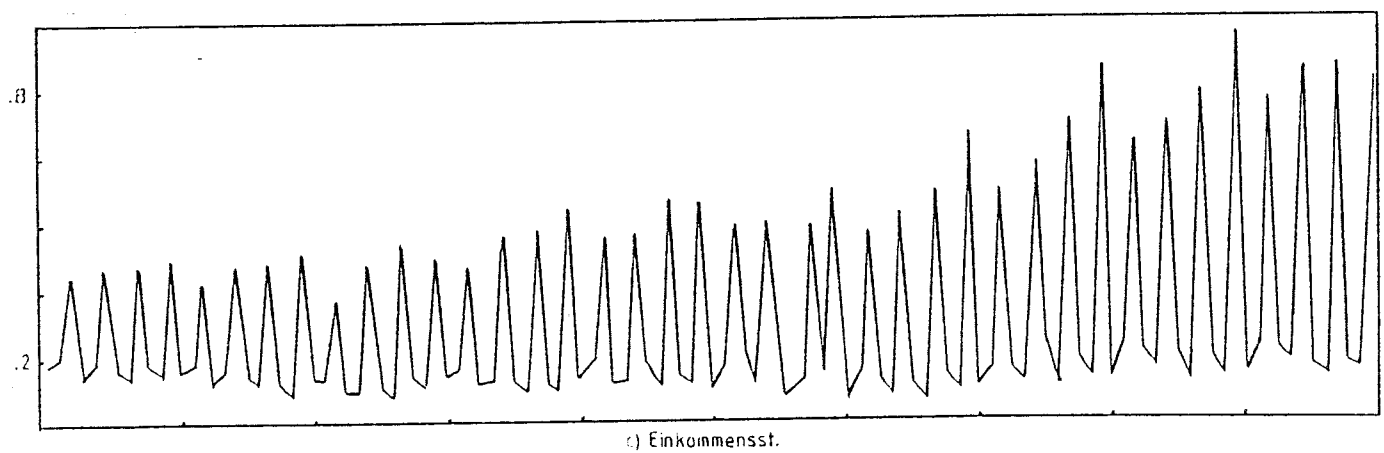
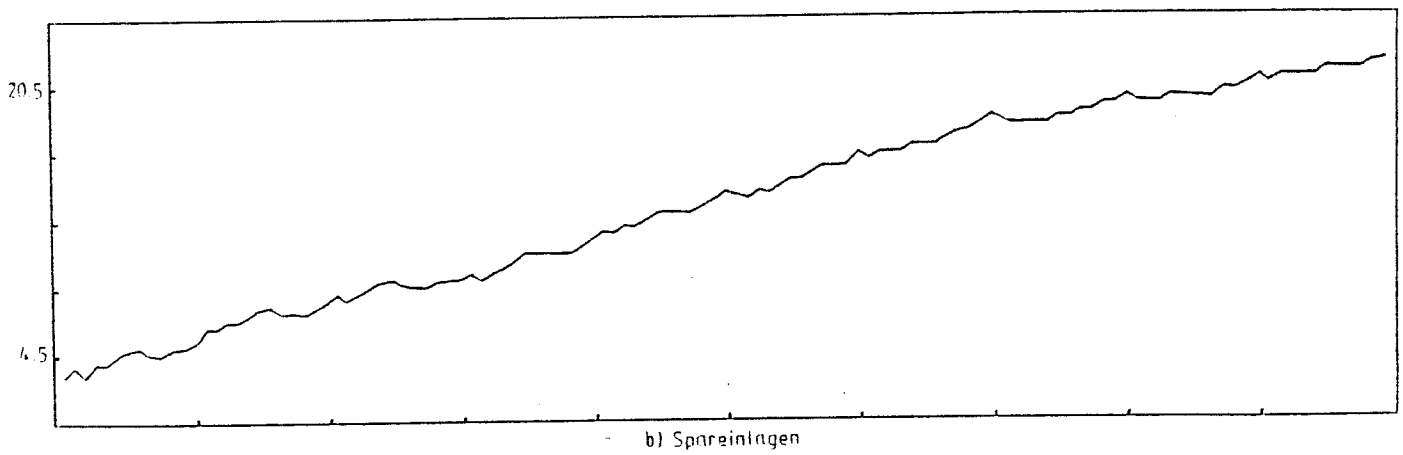
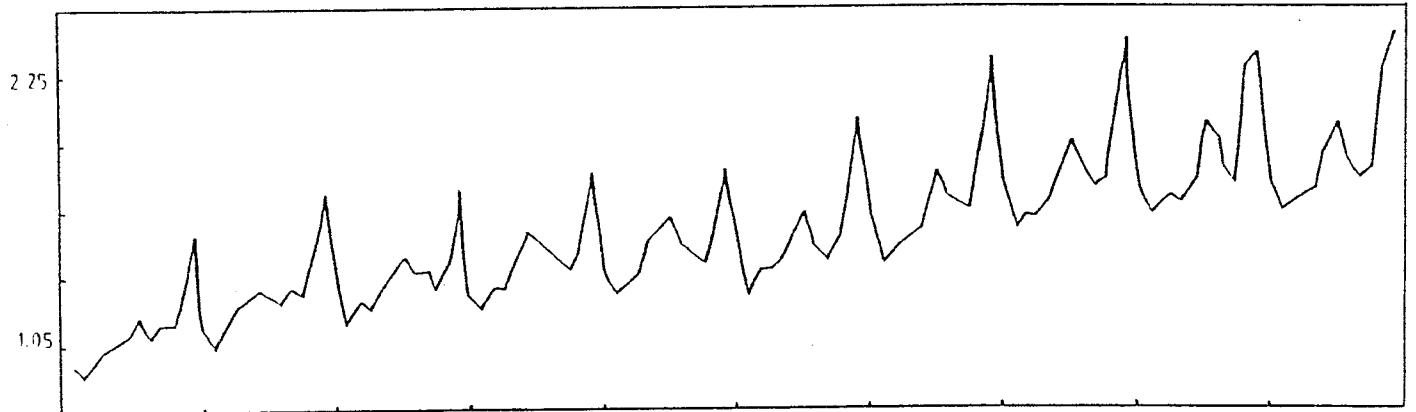
TABELLE (A.1): Beschreibung der Variablen

Var.	Bezeichnung	Einheit	WIFO-Kennz.
Z ₁	Netto-Tariflohnindex (Arbeiter), ohne Kinderbeihilfe	Apr.'45=100	YTNAOS
Z ₂	Lohn- und Gehaltssumme Brutto (Industrie), mit Sonderz.	Mill. S	YSIBIN
Z ₃	Verkaufsmenge Zigaretten (Inl.)	1000 St.	CIAZTM
Z ₄	Export Waren insgesamt (laut Handelsstatistik)	Mill. S	XZATTN
Z ₅	Import Waren insgesamt (laut Handelsstatistik)	Mill. S	MZATTN
Z ₆	Goldbestand Notenbank (gemünzt und ungemünzt)	Mill. S.	FNGOMN
Z ₇	Nationalbankbilanz Devisen und Valuten	Mill. S	FNDVNM
Z ₈	Spareinlagen bei Sparkassen	Mill. S.	FSPSPN
Z ₉	Gewerbesteueraufkommen	Mill. S.	GSGEMN
Z ₁₀	Vermögenssteueraufkommen	Mill. S.	GSVEMN
Z ₁₁	Einkommenssteueraufkommen	Mill. S.	GSEIMN
Z ₁₂	Sachgüterproduktion Erdölind.	1976 = 100	SIB04R
Z ₁₃	Sachgüterproduktion Glasind.	1976 = 100	SIF04R
Z ₁₄	Sachgüterproduktion Elektroind.	1976 = 100	SIX04R

Allen Reihen ist eine mehr oder weniger deutlicher Trend gemeinsam, mit zum Teil deutlichen und im Zeitablauf zunehmenden Saisonausschlägen (Lohnsumme und Einkommenssteuer).

Schaubild (A.2) gibt für den gleichen Zeitraum drei weitere typische Kurvenverläufe wieder. Gemeinsam ist hier ebenfalls wieder ein mehr oder weniger deutlicher Trend, überlagert von Saison-

SCHAUBILD (A.1): Ausgewählte charakteristische Kurvenverläufe



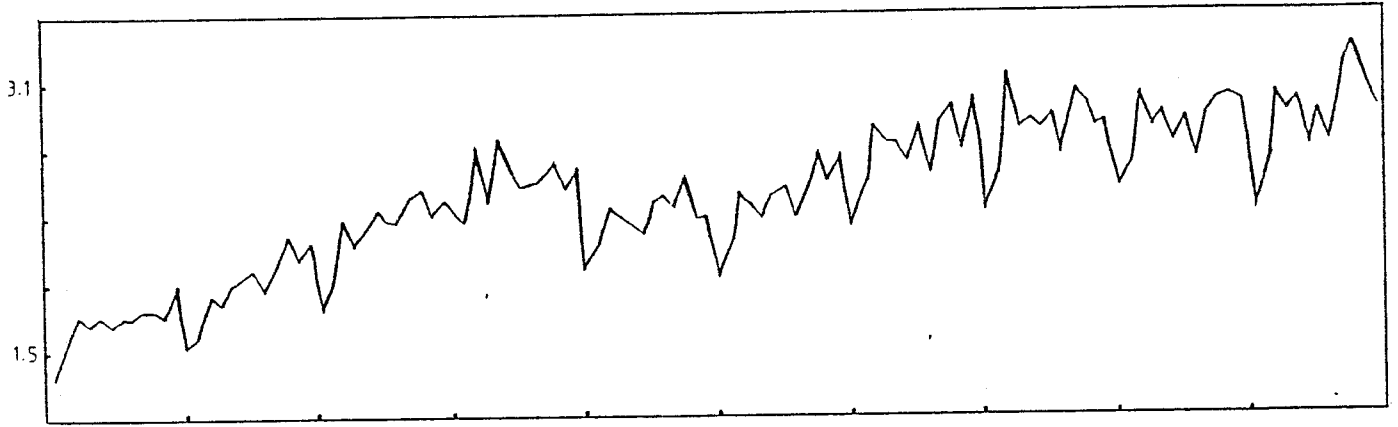
schwankungen und Konjunkturzyklen.

Tabelle (A.2) reproduziert die Ergebnisse von Regressionen (über den maximalen Beobachtungszeitraum) der Variablen aus Tabelle (A.1) gegen linearen und quadratischen Trend. Wegen stark schwankender Größenordnungen der Variablen und daher schlecht vergleichbarer Regressionskoeffizienten sind allein die resultierenden t-Werte explizit angegeben. Diese sind natürlich strikt als deskriptive Statistiken zu interpretieren und zeigen aus dieser Sicht ebenfalls die Trendabhängigkeit der hier betrachteten Variablen auf (nach Definition (2.20) und (4.17) ist Trend allein über die

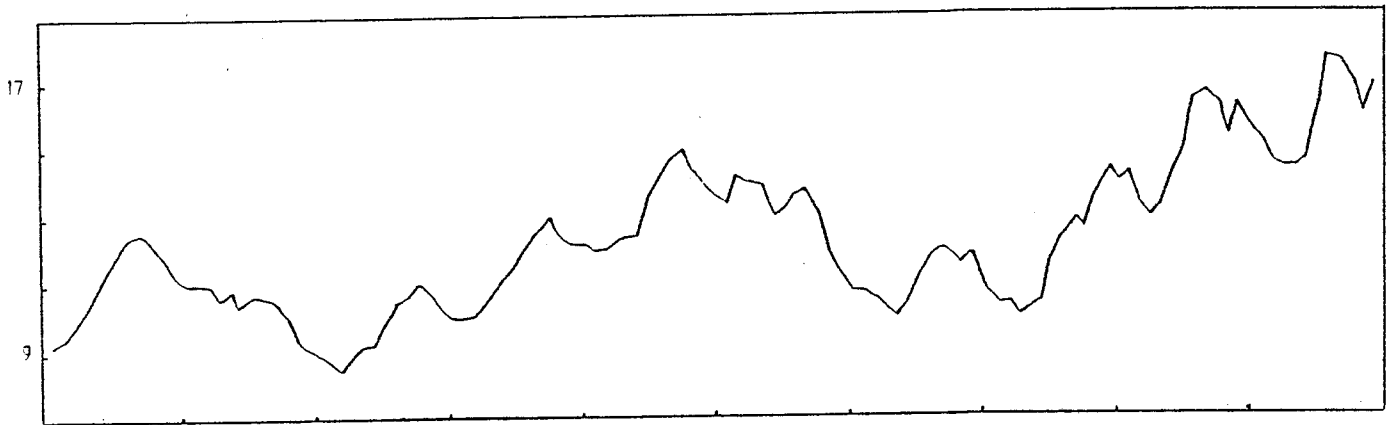
TABELLE (A.2): t-Werte bei Regression gegen linearen quadratischen Trend

Variable	Const.	T	T ²	R ²
Tariflohnindex	75.2	8.5	21.1	.98
Lohnsumme	12.3	9.7	.8	.85
Verk. Zig.	30.0	15.7	-4.6	.87
Export	19.3	3.0	15.1	.95
Import	17.0	4.7	15.4	.96
Goldbestand	-3.2	20.2	-15.9	.66
Devisen	15.0	2.6	3.7	.68
Spareinlagen	10.7	25.7	41.6	.996
Gewerbesteuer	2.1	3.6	-2.5	.08
Vermögenssteuer	1.1	1.9	.1	.18
Einkommensst.	1.7	2.1	-.3	.15
Output Erdölind.	15.6	14.7	-9.2	.67
Output Glasind.	29.3	-8.6	21.6	.90
Output Elektroind.	8.3	2.6	16.3	.95

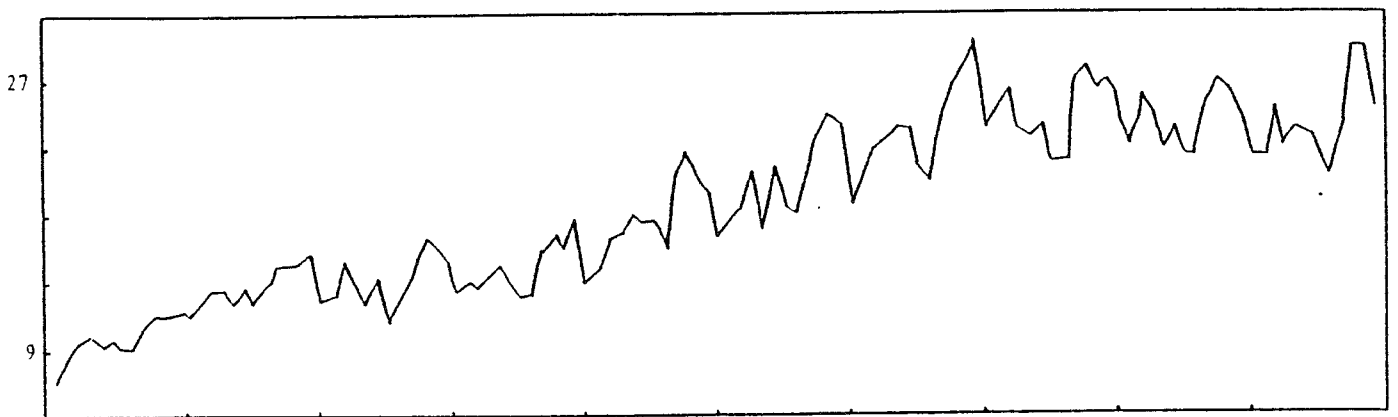
SCHAUBILD (A.2): Einige weitere ausgewählte Kurvenverläufe



a) Exporte



b) Devisen



c) Output Elektroind.

kumulierten Quadratsummen definiert; bei Regression dieser Größen gegen quadratischen und kubischen Trend ergaben sich ebenfalls stets signifikante Abhängigkeiten).

Die für das gegenwärtige Untersuchungsziel eigentlich relevanten Daten befinden sich in Tabelle (A.3). Hier werden die $K(K+1)/2 = 120$ Korrelationskoeffizienten (im Sinn der Bravais-Pearson Definition) der Zeitreihen für wachsendes T verfolgt. Die normierten Kreuzprodukte selbst werden leider vom IAS-SYSTEM nicht ausgegeben, aber man macht sich leicht klar, daß bei einer Konstanten in der Regression zur Verhinderung einer singulären Grenzmatrix der normierten Kreuzprodukte (siehe Beispiel (4.18)) notwendig ist, daß die empirischen Korrelationskoeffizienten je zweier exogener Variablen nicht gegen 1 streben. Diese Bedingung ist zum Beispiel verletzt bei Variablen mit linearem Trend und beschränkten Abweichungen vom Trend. Zusätzlich sind daher in der ersten Spalte von Tabelle (A.3) die Korrelationen mit der Variablen $z_t = t$ ($t=1, \dots, T$) angegeben.

Wie man aus dieser ersten Spalte sieht, nimmt die Korrelation mit einer linearen Trendvariablen für fast alle z_i im Zeitablauf zu. Eine Ausnahme ist etwa z_6 (Goldbestand der Nationalbank). Diese Zunahme flacht aber mit wachsendem T ab und läßt die Hypothese eines von 1 verschiedenen Grenzwertes durchaus plausibel erscheinen. Für z_8 = Spareinlagen, die für kleine T eine fast perfekte Korrelation mit einem linearen Trend aufweisen (siehe auch Schaubild(A.1b)), nimmt die Korrelation mit wachsendem T sogar ab. Für die Steueraufkommensvariablen z_9 , z_{10} und z_{11} bleibt die Korrelation während des gesamten Beobachtungszeitraums unter 0.5. Nur für

TABELLE (A.3): Korrelationskoeffizienten der Variablen für verschiedene T

t	z ₁	z ₂	z ₃	z ₄	z ₅	z ₆	z ₇	z ₈	z ₉	z ₁₀	z ₁₁	z ₁₂	z ₁₃	
a) T = 120														
z ₁	.93	1												
z ₂	.84	.83	1											
z ₃	.80	.75	.78	1										
z ₄	.87	.86	.79	.82	1									
z ₅	.88	.88	.83	.76	.93	1								
z ₆	.97	.92	.81	.76	.81	.84	1							
z ₇	.69	.63	.60	.61	.50	.47	.69	1						
z ₈	1.00	.94	.83	.78	.85	.87	.98	.69	1					
z ₉	.26	.28	.20	.15	.17	.19	.25	.21	.26	1				
z ₁₀	.38	.37	.29	.27	.28	.29	.37	.30	.38	.96	1			
z ₁₁	.30	.23	.37	.25	.32	.24	.26	.25	.26	-.36	-.33	1		
z ₁₂	.21	.32	.28	.09	.15	.24	.26	.28	.22	-.00	-.08	.16	1	
z ₁₃	.05	.19	.09	.09	.19	.16	.15	.11	.06	.12	.14	-.02	.33	1
z ₁₄	.92	.88	.81	.73	.87	.89	.89	.55	.92	.20	.33	.30	.08	.08
b) T = 240														
z ₁	.96	1												
z ₂	.90	.92	1											
z ₃	.91	.89	.88	1										
z ₄	.92	.96	.92	.89	1									
z ₅	.94	.97	.93	.88	.97	1								
z ₆	.84	.69	.67	.74	.64	.66	1							
z ₇	.69	.76	.74	.68	.74	.77	.36	1						
z ₈	1.00	.97	.91	.90	.93	.95	.82	.70	1					
z ₉	.29	.28	.31	.22	.22	.24	.26	.21	.29	1				
z ₁₀	.36	.35	.37	.29	.30	.31	.29	.26	.36	.96	1			
z ₁₁	.38	.38	.40	.41	.44	.40	.28	.37	.38	-.31	-.30	1		
z ₁₂	.89	.92	.88	.79	.88	.91	.64	.69	.90	.24	.29	.38	1	
z ₁₃	.82	.90	.84	.78	.89	.88	.51	.73	.84	.27	.35	.31	.86	1
z ₁₄	.93	.96	.92	.85	.97	.97	.63	.77	.94	.25	.32	.43	.90	.90
c) T = 348														
z ₁	.96	1												
z ₂	.92	.92	1											
z ₃	.93	.90	.89	1										
z ₄	.95	.97	.91	.89	1									
z ₅	.96	.98	.92	.89	.99	1								
z ₆	.64	.47	.50	.62	.49	.50	1							
z ₇	.81	.83	.79	.77	.80	.80	.33	1						
z ₈	.99	.98	.90	.90	.97	.97	.60	.81	1					
z ₉	.25	.23	.29	.20	.16	.20	.20	.20	.22	1				
z ₁₀	.43	.42	.45	.36	.36	.39	.25	.34	.42	.93	1			
z ₁₁	.39	.38	.38	.41	.45	.40	.23	.37	.38	-.35	-.31	1		
z ₁₂	.77	.74	.76	.74	.70	.72	.49	.65	.71	.22	.30	.35	1	
z ₁₃	.88	.90	.83	.80	.93	.93	.46	.71	.92	.17	.38	.34	.55	1
z ₁₄	.95	.97	.92	.87	.99	.98	.48	.80	.97	.18	.37	.44	.71	.93

z_{14} = Output Elektroindustrie legt die Tabelle deutlich den Schluß nahe, bei Koexistenz von $z_t=t$ und z_{14} unter den exogenen Variablen eines Modells die Annahmen der vorliegenden Arbeit als unplausibel abzulehnen.

Von Interesse sind natürlich auch die Korrelationen der z_i ($i=1, \dots, 14$) untereinander. Auch hier lassen sich Variablenpaare finden, deren empirische Korrelationen zielstrebig die 1 anzusteuern scheinen und die daher, falls gemeinsam unter den exogenen Variablen eines der hier betrachteten Modelle vertreten, unsere Voraussetzungen sehr unplausibel machen würden. Beispiele sind etwa die Paare Import-Export oder Devisen-Spareinlagen. Auf der anderen Seite finden sich aber auch genügend Variablenpaare, deren Korrelationen keine Tendenz gegen 1 aufweisen, teilweise sogar abnehmen oder negativ sind.

Natürlich ist die fehlende Konvergenz gegen 1 der empirischen Korrelationskoeffizienten nur eine notwendige, keine hinreichende Bedingung für eine reguläre Grenzmatrix der normierten Kreuzprodukte. In einer abschließenden Tabelle werden daher einige Kennziffern einer typischen Regressorenmatrix Z im Zeitablauf verfolgt. Die dazu willkürlich ausgewählten Variablen sind (i) eine Konstante, (ii) ein linearer Trend, (iii) z_1 =Tariflohnindex, (iv) z_4 =Exporte, und (v) z_{11} =Einkommenssteueraufkommen. Tabelle (A.4) gibt den größten und kleinsten Eigenwert von $Z'Z$ (λ_{\max} bzw. λ_{\min}) und die Determinante von $D_t^{-1} Z'Z D_T^{-1}$ für verschiedene T an, bis zum maximalen vom IAS-SYSTEM für diesen Zweck akzeptierten Wert $T=204$.

Die Tabelle zeigt zunächst, daß der größte Eigenwert von $Z'Z$ schneller gegen ∞ strebt als der Stichprobenumfang, ein Kenn-

TABELLE (A.4): Einige Kennziffern von $Z'Z$ für verschiedene T

Zeitraum	T	$\lambda_{\max}(Z'Z)$	$\lambda_{\min}(Z'Z)$	$\det(D_T^{-1}Z'ZD_T^{-1})$
Jan.'54-Dez.'58	60	126710	.008	$16 \cdot 10^{-5}$
Jan.'54-Dez.'62	108	583294	.023	$26 \cdot 10^{-5}$
Jan.'54-Dez.'66	156	1592262	.060	$44 \cdot 10^{-5}$
Jan.'54-Dez.'70	204	3384049	.124	$120 \cdot 10^{-5}$

zeichen für Trend in der Regressorenmatrix. Zusätzlich nehmen aber auch der kleinste Eigenwert und insbesondere die Determinante der normierte Matrix im Zeitablauf zu. Zwar ist die Determinante absolut sehr klein, eine Folge des hohen Multikollinearitätsgrads der Regressorenmatrix, aber eine Tendenz zur Singularität läßt sich nicht erkennen.

Als Resümé läßt sich daher festhalten, daß die in dieser Arbeit getroffenen Annahmen zum Grenzverhalten ökonometrischer Regressorenmatrizen in vielen Anwendungsfällen zumindest nicht un-
plausibel erscheinen.

b) RELATIVE EFFIZIENZ DES ERSTE-DIFFERENZEN SCHÄTZERS
BEI WACHSENDER STÖRGRÖSSENKORRELATION

Der Erste-Differenzen-Schätzer (FD für "First Differences; siehe Abschnitt 3a) der Steigungskoeffizienten eines linearen Regressionsmodells dient bei vermuteter hoher positiver Korrelation der Störgrößen oft als praktikable Approximation der allein theoretisch interessanten Verallgemeinerten Kleinst-Quadrate Methode (GLS). Diese Vorgangsweise soll hier auf ihre Gültigkeit untersucht und in den Kontext trendbehafteter exogener Variablen gestellt werden. Entgegen KADIYALA (1968) und MAESHIRO (1976) erweist sich FD für stationäre AR(1)-Störgrößen dabei in der Tat als asymptotisch BLUE, gegeben einen festen Stichprobenumfang T und wachsende Störgrößenkorrelation. Dies gilt für beliebige Regressorenmatrizen und wurde auch von MAESHIRO (1980b) zugestanden, aber ein formeller Beweis fehlt bisher (für den Spezialfall des Modells aus Abschnitt 3a folgt dieses Resultat unmittelbar aus 3.15). Andererseits ist diese Eigenschaft von FD auf AR(1) Störgrößenprozesse beschränkt und gilt nicht mehr notwendig in Situationen, in denen die Korrelation benachbarter Störgrößen zwar gegen Eins strebt, aber die Störgrößen einem anderen als AR(1)-Prozess folgen.

Das im weiteren zugrundegelegte Modell ist

$$y = X\beta + u, \quad (A.1)$$

mit einem $T \times 1$ -Beobachtungsvektor y der abhängigen und einer $T \times K$ -Beobachtungsmatrix der unabhängigen Variablen ($T \times K$, $\text{Rang}(X)=K$). Der $T \times 1$ -Störgrößenvektor u habe Erwartungswert 0 und eine endliche Varianz-Kovarianzmatrix V mit $\text{Rang}(V)=T$. Im Gegensatz zu Abschnitt 2a beschränken wir uns hier auf nichtstochastische Regressoren, lassen aber korrelierte Störgrößen zu.

Die GLS-Schätzung $\hat{\beta}^{\text{GLS}}$ für β ist gegeben als OLS, angewandt auf

$$P y = P X \beta + P u \quad , \quad (\text{A.2})$$

mit einer $T \times T$ -Matrix P , die $PVP' = \sigma^2 I$ erfüllt. Wie man sich leicht klarmacht, läuft dies gerade auf die Formel (3.3) hinaus. Ähnlich ist $\hat{\beta}^{\text{FD}}$ definiert als OLS, angewandt auf

$$D y = D X \beta + D u \quad . \quad (\text{A.3})$$

Dabei ist

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \cdot & \ddots & \cdot & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \quad (\text{A.4})$$

die $(T-1) \times T$ Differenzenmatrix.

Aus (A.2) und (A.3) folgt unmittelbar

$$\hat{\beta}^{\text{GLS}} = ((PX)'(PX))^{-1}(PX)'Py \quad (\text{A.5})$$

und

$$\hat{\beta}^{\text{FD}} = ((DX)'(DX))^{-1}(DX)'Dy \quad . \quad (\text{A.6})$$

Die Varianz-Kovarianzmatrizen dieser Schätzfunktionen sind

$$\hat{V}^{\text{GLS}} := \sigma^2 ((\text{PX})'(\text{PX}))^{-1} \quad (\text{A.7})$$

und

$$\hat{V}^{\text{FD}} := ((\text{DX})'(\text{DX}))^{-1} (\text{DX})' \text{DVD}' \text{DX} ((\text{DX})'(\text{DX}))^{-1} . \quad (\text{A.8})$$

Bei inhomogener Regression, d.h. einer ersten Spalte von X identisch gleich Eins, wird die erste Spalte von DX gleich Null und entfällt vor der Anwendung von OLS auf (A.3). Sowohl $\hat{\beta}^{\text{FD}}$ als auch \hat{V}^{FD} beziehen sich dann nur auf die Steigungskoeffizienten des Modells.

Wir betrachten zunächst einen AR(1)-Störgrößenprozess, d.h.

$$u_t = \rho u_{t-1} + \varepsilon_t . \quad (\text{A.9})$$

Die Notation folgt derjenigen in Abschnitt 3a. Unter der zusätzlichen Annahme der Stationarität entspricht ρ gerade der Korrelation je zweier benachbarter Störgrößen, und gilt

$$\sigma_u^2 = \sigma_\varepsilon^2 / (1 - \rho^2) . \quad (\text{A.10})$$

σ_ε^2 wird dabei im weiteren konstant gehalten.

SATZ (A.1): Für $\rho \rightarrow 1$ konvergieren sowohl \hat{V}^{GLS} als auch \hat{V}^{FD} gegen die gleiche endliche $K \times K$ -Matrix Q (homogene Regression),

oder

für $\rho \rightarrow 1$ konvergieren sowohl \hat{V}^{FD} als auch die rechte untere $(K-1) \times (K-1)$ -Teilmatrix von \hat{V}^{GLS} gegen die gleiche $(K-1) \times (K-1)$ -Matrix Q_1 (inhomogene Regression).

BEWEIS: Als GLS-Transformationsmatrix wählen wir zunächst wie in Abschnitt 3a

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$$

Daraus folgt unmittelbar $\lim_{\rho \rightarrow 1} P = (0:D)'$ und $\lim_{\rho \rightarrow 1} (PX)'(PX) = (DX)'(DX)$.

Bei homogener Regression (keine Konstante) ist $(DX)'(DX)$ regulär, und damit wegen der angenommenen Konstanz von σ_ε^2

$$\begin{aligned} \lim_{\rho \rightarrow 1} \hat{V}^{\text{GLS}} &= \lim_{\rho \rightarrow 1} \sigma_\varepsilon^2 ((PX)'(PX))^{-1} \\ &= \sigma_\varepsilon^2 ((DX)'(DX))^{-1} \\ &= Q \end{aligned} \quad (\text{A.11})$$

mit einer regulären und endlichen Matrix Q . Andererseits folgt $\lim_{\rho \rightarrow 1} \hat{V}^{\text{FD}}$ sofort aus (A.8),

$$DVD' = \frac{\sigma_\varepsilon^2}{1+\rho} \begin{bmatrix} 2 & \rho-1 & \dots & \rho^{T-3}(\rho-1) \\ \rho-1 & 2 & \dots & \rho^{T-4}(\rho-1) \\ \rho(\rho-1) & -1 & \dots & \rho^{T-5}(\rho-1) \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-3}(\rho-1) & \rho^{T-4}(\rho-1) \dots & & 2 \end{bmatrix}$$

und $\lim_{\rho \rightarrow 1} DVD' = \sigma_\varepsilon^2 I$, und der erste Teil von Satz (A.1) ist bewiesen.

Zum Beweis des zweiten Teils schreibe man (A.1) als

$$y = (1:X_1) \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} + u, \quad (\text{A.12})$$

wobei 1 einen $T \times 1$ -Vektor von Einsen und α das Absolutglied der Gleichung bezeichnen soll. $\hat{V}_1^{\text{FD}} = \text{cov}(\hat{\beta}_1^{\text{FD}})$ ist dann weiterhin durch (A.8) gegeben (nach Vertauschung von X und X_1), während \hat{V}_1^{GLS} nun die untere rechte $(K-1) \times (K-1)$ -Teilmatrix von

$$\hat{V}^{\text{GLS}} = \sigma_\varepsilon^2 \begin{bmatrix} (P1)'P1 & (P1)'PX_1 \\ (PX_1)'P1 & (PX_1)'PX_1 \end{bmatrix}^{-1}$$

ist, d.h. (siehe etwa THEIL, 1971, S. 18)

$$\hat{V}_1^{\text{GLS}} = \sigma_\varepsilon^2 \left((PX_1)'PX_1 - (PX_1)'P1((P1)'(P1))^{-1}(P1)'PX_1 \right)^{-1}. \quad (\text{A.13})$$

Ferner hat

$$(P1)'(P1) = 1 - \rho^2 + (T-1)(1-\rho)^2 \quad (\text{A.14})$$

eine einfache Nullstelle für $\rho=1$, und

$$(P1)(P1)' = \begin{bmatrix} (1-\rho)(1+\rho) & \sqrt{1-\rho^2}(1-\rho) & \dots & \sqrt{1-\rho^2}(1-\rho) \\ \sqrt{1-\rho^2}(1-\rho) & (1-\rho)^2 & & (1-\rho)^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{1-\rho^2}(1-\rho) & (1-\rho)^2 & \dots & (1-\rho)^2 \end{bmatrix}$$

Daher ist $\lim_{\rho \rightarrow 1} P1((P1)'(P1))^{-1}(P1)'$ überall außer in der oberen rechten Ecke gleich Null. Wegen $\lim_{\rho \rightarrow 1} PX = (0:DX)'$ verschwindet damit der zweite Term in (A.13) für $\rho \rightarrow 1$. Das impliziert aber

$$\lim_{\rho \rightarrow 1} \hat{V}_1^{\text{GLS}} = \sigma_\varepsilon^2 ((DX_1)(DX_1))^{-1} = \lim_{\rho \rightarrow 1} \hat{V}_1^{\text{FD}},$$

q.e.d.

Satz (A.1) liefert eine gewisse Rechtfertigung für die in der ökonometrischen Praxis häufig beobachtete Vorgehensweise, bei vermuteter hoher positiver Korrelation der Störgrößen GLS durch FD zu approximieren. Dies gilt jedoch nur für AR(1) Störgrößenprozesse. Geht die Korrelation je zweier benachbarter Störgrößen gegen Eins, aber ohne daß die Störgrößen einem AR(1) Prozeß folgen, ist auch FD nicht mehr notwendig asymptotisch effizient. Man betrachte etwa einen stationären AR(2) Störgrößenprozeß mit

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t. \quad (\text{A.15})$$

Hier ist die Korrelation je zwei benachbarter Störgrößen gegeben durch

$$r = \rho_1 / (1 - \rho_2) \quad . \quad (\text{A.16})$$

Ferner sei X eine $T \times 1$ -Matrix mit $X = (1, 0, \dots, 0)'$, und $\rho_1 = \rho_2$.

Dann gilt $\lim_{\rho_1 \rightarrow 1/2} r = 1$, aber man rechnet leicht nach, daß $\lim_{\rho_1 \rightarrow 1/2} \text{var}(\hat{\beta}^{\text{GLS}}) = \sigma_\varepsilon^2$ und $\lim_{\rho_1 \rightarrow 1/2} \text{var}(\hat{\beta}^{\text{FD}}) = 4 \sigma_\varepsilon^2$.

Sogar der Fall einer für $r \rightarrow 1$ verschwindenden relativen Effizienz von FD ist nicht ausgeschlossen. Dazu sei X wie bisher, und $\rho_2 = 1 - \rho_1 - \rho_1^2$. Wegen (A.16) gilt dann $\lim_{\rho_1 \rightarrow 0} r = 1$. Auf der anderen Seite prüft man aber leicht nach, daß $\lim_{\rho_1 \rightarrow 1} \text{var}(\hat{\beta}^{\text{GLS}}) = \sigma_\varepsilon^2$, während $\text{var}(\hat{\beta}^{\text{FD}})$ für $\rho_1 \rightarrow 0$ über alle Grenzen wächst (man beachte dazu, daß σ_u^2 wegen der Konstanz von σ_ε^2 und $\sigma_u^2 = \sigma_\varepsilon^2 / (1 - \rho_1 - \rho_2 - \rho_1^2 \rho_2 / (1 - \rho_2))$ bei wachsender Störgrößenkorrelation ebenfalls zunehmen muß).

Seien nun die Störgrößen des Modells (A.1) Ausschnitt eines beliebigen stationären AR(p)-Prozesses, d.h. insbesondere

$$u_t = \rho_1 u_{t-1} + \dots + \rho_p u_{t-p} + \varepsilon_t \quad (\text{A.17})$$

mit $\rho_p \neq 0$. Alle Nullstellen des charakteristischen Polynoms von (A.17) seien betragsmäßig größer Eins, und r_s bezeichne die Korrelation zwischen u_t und u_{t+s} . Derartige Störgrößenprozesse wurden bereits in Abschnitt 3b untersucht. Wegen der Stationarität des Prozesses hat man ferner

$$\sigma_u^2 = \sigma_\varepsilon^2 / \left(1 - \sum_{i,j=1}^p \rho_i \rho_j r_{i-j}\right) \quad (\text{A.18})$$

und

$$r_1 \rightarrow 1 = \sum_{i=1}^p \rho_i \rightarrow 1 \quad (\text{A.19})$$

SATZ (A.2): Für $r_1 \rightarrow 1$ konvergieren sowohl \hat{V}^{GLS} als auch \hat{V}^{FD} (\hat{V}_1^{GLS} und \hat{V}_1^{FD}) gegen die gleiche endliche und reguläre Grenzmatrix Q (Q_1), unabhängig von X , dann und nur dann wenn

$$\rho_1 \rightarrow 1, \rho_i \rightarrow 0 \quad (i \neq 1) \quad (\text{A.20})$$

BEWEIS: \hat{V}^{GLS} und \hat{V}^{FD} (bzw. \hat{V}_1^{GLS} und \hat{V}_1^{FD}) sind stetige Funktionen von ρ_i ($i=1, \dots, p$) und σ_ε^2 . Daher gilt

$$\begin{aligned} \lim_{\rho_1 \rightarrow 1; \rho_i \rightarrow 0 (i \neq 1)} \hat{V}(\rho_1, \dots, \rho_p, \sigma_\varepsilon^2) \\ = \lim_{\rho_1 \rightarrow 1} \hat{V}(\rho_1, 0, \dots, 0, \sigma_\varepsilon^2) \end{aligned} \quad (\text{A.21})$$

mit $\hat{V} = \hat{V}^{\text{GLS}}, \hat{V}^{\text{FD}}, \hat{V}_1^{\text{GLS}}, \hat{V}_1^{\text{FD}}$, und (A.20) ist hinreichend für Satz (A.2) wegen Satz (A.1).

Zum Beweis der Notwendigkeit gelte $r_1 \rightarrow 1$, aber $\rho_1 \rightarrow c \neq 1$. Es reicht zu zeigen, daß immer ein X existiert mit $\lim_{r_1 \rightarrow 1} \hat{V}^{\text{GLS}} \neq \lim_{r_1 \rightarrow 1} \hat{V}^{\text{FD}}$. Der Einfachheit halber sei $K=1$ und $X = (1, \dots, 0)'$ wie bisher. Dann gilt $\text{var}(\hat{\beta}^{\text{GLS}}) = \sigma_\varepsilon^2$ unabhängig von den ρ_i , und daher $\lim_{r_1 \rightarrow 1} \text{var}(\hat{\beta}^{\text{GLS}}) = \sigma_\varepsilon^2$ a fortiori.

Andererseits gilt hier

$$\begin{aligned} \hat{\beta}^{\text{FD}} - \beta &= u_1 - u_2 \\ &= u_1 - \rho_1 u_1 - \rho_2 u_0 - \dots - \rho_p u_{2-p} - \varepsilon_2 \\ &= (1-\rho_1) \varepsilon_1 - \varepsilon_2 + z. \end{aligned} \quad (\text{A.22})$$

Dabei ist z eine Linearkombination von u_i 's mit $i < 1$ und daher unkorreliert mit $(1-\rho_1)\varepsilon_1 - \varepsilon_2$. Daraus folgt aber

$$\begin{aligned} \text{var}(\hat{\beta}^{\text{FD}}) &= ((1-\rho_1)^2 + 1)\sigma_\varepsilon^2 + \sigma_z^2 \\ &\geq ((1-\rho_1)^2 + 1)\sigma_\varepsilon^2. \end{aligned} \quad (\text{A.23})$$

Da ex Annahme $\lim_{r_1 \rightarrow 1} \rho_1 \neq 1$, liefert (A.23) unmittelbar die Behauptung, q.e.d.

Die in den Sätzen (A.1) und (A.2) gebrauchte Annahme einer konstanten Varianz der Innovationen ε_t hat vor allem beweistechnische Bedeutung und ist für das Ergebnis unerheblich. Wegen (A.18) und (A.19) folgt daraus etwa eine für $r_1 \rightarrow 1$ gegen ∞ strebende Varianz der Störgrößen selbst, aber da hier allein die Quotienten der Varianzen von GLS und FD interessieren, ist diese

Implikation hier ohne Bedeutung. Alternativ könnte man auch σ_u^2 festhalten, wie in KRÄMER (1980b).

Satz (A.2) zeigt, daß unter allen autoregressiven Störgrößenprozessen diejenigen der Ordnung Eins im wesentlichen die einzigen sind, für die man bei hoher positiver Störgrößenkorrelation und unabhängig von der konkreten Gestalt von X eine gute Approximation von GLS durch FD erwarten kann. Wie ferner in Abschnitt 3a gezeigt, ist auch dann die Güte der Approximation bei Trend in den Daten nicht unabhängig von T. Dies mag daher abschließend als Hinweis dienen, bei Trend in den Daten eine Entscheidung für FD nicht nur von der vermuteten Störgrößenkorrelation, sondern auch von der Stichprobengröße abhängig zu machen.

Literaturverzeichnis

Es werden die folgenden Abkürzungen verwendet:

JASA: Journal of the American Statistical Association
 JRSS: Journal of the Royal Statistical Society
 JOE: Journal of Econometrics
 IER: International Economic Review

Aigner, D.J., und Goldberger, A.S. (1977), Latent variables in socio-economic models, Amsterdam (North Holland).

———, Hsiao, Ch., Kapteyn, A., und Wansbaek, T. (1984), "Latent variable models in econometrics," in: Griliches, Z. und Intriligator, M.D. (Hrsg.), Handbook of Econometrics, Vol. 6, Amsterdam (North Holland).

Anderson, O.D. (1979), "Some sample autocovariance function results for a once integrated q^{th} -order moving average process," Statistica 2, 287-299.

Anderson, T.W. (1959), "On asymptotic distributions of estimates of parameters of stochastic difference equations," Annals of Mathematical Statistics, 30, 676-687.

——— (1971), The statistical analysis of time series, New York (Wiley).

——— (1976), "Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics," JRSS, B 38, 1-20.

———, und Sawa, T. (1982), "Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient," JRSS B, 44, 52-62.

- , und Taylor, J.B. (1976), "Strong consistency of least squares estimates in normal linear regression," Annals of Statistics 4, 788-790.
- , und ——— (1979), "Conditions for strong consistency of least squares estimates in linear models," Annals of Statistics 7, 484-489.
- Atkinson, S.E. (1978), "Small sample properties of simultaneous equation estimators with multicollinearity," JASA 73, 719-723.
- Basman, R.L. (1961), "A note on the exact finite sample frequency functions of GCL estimators in two leading over-identified cases," JASA 56, 619-636.
- (1963), "Remarks concerning the application of exact finite sample distribution functions of GCL estimators in econometric statistical inference," JASA 58, 943-976.
- (1974), "Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal," in Intriligator, M.D. und Kendrick, D.A. (Hrsg.), Frontiers of quantitative economics, Vol. 2, Amsterdam (North Holland).
- , Richardson, D.H., und Rohr, R.J. (1974a), "An experimental study of structural estimators and test statistics associated with dynamical econometric models," Econometrica 42, 717-730.
- , ———, und ——— (1974b), "Finite sample distributions associated with stochastic differences equations: some experimental evidence," Econometrica 42, 825-839.
- Box, G.E.P., und Jenkins, G.M. (1970), Time series analysis, San Francisco (Holden Day).
- Chan, K.H., Haya, J.C., und Ord, J.K. (1977), "A note on trend removal methods: the case of polynomial regression versus variate differencing," Econometrica 45, 737-744.
- Chipman, J.S. (1979), "Efficiency of least squares estimation of linear trend when residuals are autocorrelated," Econometrica 47, 115-128.

- Christopeit, N., und Helmes, K. (1979), "Strong consistency of generalized least squares estimators," Annals of Statistics 7, 795-800.
- Conrad, K., und Jüttner, D.J. (1973), "Recent behaviour of stock market prices in Germany and the random walk hypothesis," Kyklos 26, 576-599.
- Cragg, J.G. (1968), "Some effects of incorrect specification on small sample properties of several simultaneous equation estimators," IER 9, 63-68.
- Dhrymes, Ph. (1970), Econometrics, New York (Harper&Row).
- Dickey, D.A., und Fuller, W. (1981), "Likelihood ratio statistics for autoregressive time series with a unit root," Econometrica 49, 1057-1072.
- Drygas, H. (1973), "A note on a paper by T. Kloek concerning the consistency of variance estimation in the linear model," Econometrica 43, 175.
- (1976), "Weak and strong consistency of least squares estimators in linear models," Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete, 34, 119-127.
- Eicker, F. (1963), "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," Annals of Mathematical Statistics 34, 447-456.
- (1967), "Limit theorems for regressions with unequal and dependent errors," in: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, San Francisco (University of California Press), 59-82.
- (1982), "A moment free consistency proof for the LSEs and model identification in mixed autoregressions," in: Gupta, S.S. und Berger, J.O. (Hrsg.), Statistical decision theory and related topics III, Vol. 1, New York (Academic Press), 353-383.

- Evans, G.B.A., und Savin, N.A. (1981), "Testing for unit roots: 1," Econometrica 49, 753-779.
- Feldstein, M.S. (1974), "Errors in variables: A consistent estimator with smaller MSE in finite samples," JASA 69, 990-996.
- Freimann, K.D. (1982), Monte-Carlo Studie zur Parameterschätzung in interdependenten, dynamischen ökonomischen Gleichungssystemen bei gleichzeitigem Vorliegen eines vektorautoregressiven Prozesses der Störvariablen, Diss. Hagen.
- , und Gruber, J. (1983), "Small sample properties of three estimators in dynamic simultaneous equation models with disturbances generated by a vektorautoregressive process -A Monte Carlo study," Vortrag auf der Jahrestagung der Econometric Society, Pisa.
- Frisch, R., und Waugh, F.V. (1933), "Partial time regression as compared to individual trends," Econometrica 1, 387-401.
- Fuller, W.A. (1976), Introduction to statistical time series, New York (Wiley).
- Gänssler, P., Stobel, J., und Stute, W. (1978), "On central limit theorems for martingale triangular arrays," Acta Math. Sci. Hungaricae 31, 205-216.
- Gould, J.P., und Nelson, Ch.R. (1974), "The stochastic structure of the velocity of money," American Economic Review 64, 405-417.
- Granger, C.W., und Morgenstern, O. (1970), Predictability of stock market prices, Lexington (Heath and Company).
- , und Newbold, P. (1974), "Spurious regressions in econometrics," JOE 2, 111-120.
- Grenander, U. (1954), "On the estimation of regression coefficients in the case of an autocorrelated disturbance," Annals of Mathematical Statistics 25, 252-272.
- , und Rosenblatt, M. (1957), Statistical analysis of stationary time series, New York (Wiley).

- Griliches, Z. (1974), "Errors in variables and other unobservables," Econometrica 42, 971-978.
- Haavelmo, T. (1947), "Methods of measuring the marginal propensity to consume," JASA 42, 105-122.
- Hale, C., Mariano, R.S., und Ramage, J.G. (1980), "Finite sample analysis of misspecification in simultaneous equation models," JASA 75, 418-427.
- Hammersley, J.M., und Handscomb, D.C. (1964), Monte Carlo methods, London (Methuen).
- Hall, P., und Heyde, C.C. (1980), Martingale limit theory and its application, New York (Wiley).
- Hall, R.E. (1978), "Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence," Journal of Political Economy 86, 971-987.
- Hannan, E.J. (1960), Time series analysis, London (Methuen).
- Hansen, G. (1969), "Die Prognoseeignung von LISE- und FIND-Maximum Likelihood Schätzwerten," Statistische Hefte 10, 22-45.
- Hasza, D.P. (1980), "The asymptotic distribution of the sample autocorrelations for an integrated ARMA process," JASA 75, 349-352.
- Heiler, S. (1980), "Prediction of economic processes with linear regression part," in: Nerlove et al. (Hrsg.), Problems of time series analysis, Zürich (Birkhäuser).
- Hendry, D.F., und Harrison, R.W. (1974), "Monte Carlo methodology and the small sample behaviour of ordinary and two-stage least squares," JOE 2, 151-174.
- Hillier, G.H., Kinal, T.W., und Srivastava, V.K. (1984), "On the moments of ordinary least squares and instrumental variables estimators in a general structural equation," Econometrica 52, 185-202.
- Huber, P. (1973), "Robust regression: asymptotics, conjectures and Monte Carlo," Annals of Statistics 1, 799-821.

- Jöckel, K.-H. (1982), "Iterierte Aitken-Schätzer," Allgemeines Statistisches Archiv 66, 361-357.
- Johnston, J. (1972), Econometric methods (2. Aufl.), New York (McGraw-Hill).
- Jöreskog, K.G., und Goldberger, A.S. (1975), "Estimation of a model with multiple indicators and multiple causes of a single latent variable," JASA 70, 631-639.
- Kadane, J.B. (1971), "Comparison of k-Class estimators when disturbances are small," Econometrica 39, 723-738.
- Kadiyala, K.R. (1968), "A transformation used to circumvent the problem of autocorrelation," Econometrica 36, 93-96.
- Kanter, M., und Steiger, W.L. (1974), "Regression and autoregression with infinite variance," Adv. Appl. Prob. 6, 768-783.
- Ketellapper, R.H. (1982), "The relevance of large sample properties for the errors-in-variables model: a Monte Carlo study," Communications in Statistics B 11, 625-634.
- (1983), "On estimating parameters in a simple linear errors-in-variables model," Technometrics 25, 43-47.
- Kloek, T. (1972), "Note on consistent estimation of the variance of the disturbances in the linear model," Econometrica 40, 911-912.
- Krämer, W. (1979), "Eine Alternative zur Methode der Gleitenden Durchschnitte bei der Trendschätzung mit trendabhängiger Saisonkomponente," Allgemeines Statistisches Archiv 63, 386-393.
- (1980a), Eine Rehabilitation der Gewöhnlichen Kleinst-Quadrat Methode als Schätzverfahren in der Ökonometrie, Frankfurt (Haag&Herchen).
- (1980b), "Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors," JASA 75, 1005-1009.
- (1981), "Ordinary least squares estimation of simultaneous equation systems with trended data," Vortrag auf der Jahrestagung der Econometric Society, Amsterdam.

- (1982), "Note on estimating linear trend when residuals are autocorrelated," Econometrica 50, 1065-1067.
- 1983a), "The relative efficiency of the first difference estimator in linear models with correlated errors," Methods of Operations Research 47, 65-74.
- (1983b), "Kleinst-Quadrate Schätzung ökonomischer Error-Modelle mit Trend in exogenen Variablen," Allgemeines Statistisches Archiv 67, 252-259.
- (1983c), "Some Monte Carlo evidence on the relative importance of models and methods in econometrics," Sankya B, 60-68.
- (1984a), "Some consequences of trend for simultaneous equation estimation," Economics letters 14, 23-30.
- (1984b), "Asymptotische Verteilung einiger Schätzverfahren bei Trend und Fehlern in den Variablen," erscheint in Metrika.
- Ladd, G.W. (1956), "Effects of shocks and errors in estimation: an empirical comparison," Journal of Farm Economics 38, 485-495.
- Laha, R.G., und Rohatgi, V.K. (1979), Probability theory, New York (Wiley)
- Lai, T.L., Robbins, H., und Wei, C.Z. (1978), "Strong consistency of least squares estimates in linear regression," Proc. Nat. Acad. Sci. USA 75, 3034-3036.
- , und Wei, C.Z. (1982), "Least squares estimation in stochastic regression models with applications to identification and control of dynamic systems," Annals of Statistics 10, 154-166.
- de Leeuw, J., Keller, W.J., und Wansbaek, T. (Hrsg.) (1983), Interfaces between econometrics and psychometrics (Annals of applied econometrics 1983-2), Amsterdam (North Holland).
- Madansky, A. (1959), "The fitting of straight lines when both variables are subject to error," JASA 54, 173-205.

- Maddala, G.S. (1974), "Some small sample evidence on tests of significance in simultaneous equation models," Econometrica 42, 841-851.
- Maeshiro, A. (1976), "Autoregressive transformation, trended independent variables and autocorrelated disturbance terms," Review of Economics and Statistics 58, 497-500.
- (1979), "On the retention of the first observations in serial correlation adjustment of regression models," IER 20, 259-265.
- (1980a), "New evidence on the small sample properties of estimators of SUR models with autocorrelated disturbances," JOE 12, 177-187.
- (1980b), "Autocorrelation and trended explanatory variables: a reply," Review of Economics and Statistics 62, 487-489.
- Malinvaud, E. (1981), Statistical methods of econometrics (3. Aufl.), Amsterdam (North Holland).
- Mariano, R.S. (1982), "Analytical small sample distribution theory in econometrics: the simultaneous-equation case," IER 23, 503-533.
- , und Ramage, J.G. (1982), "Specification error analysis in linear simultaneous systems," in: Basman, R.L. und Rhodes, G.F. (Hrsg.): Advances in econometrics, Band 1, London (JAI-Press)
- Mikhail, W.M. (1972), "Simulating small-sample properties of econometric estimators," JASA 67, 620-624.
- (1975), "A comparative Monte Carlo study of the properties of econometric estimators," JASA 70, 94-104.
- Neiswanger, W.A., und Yancey, T.A. (1959), "Parameter estimates and autonomous growth," JASA 54, 389-402.
- Nerlove, M., Grether, D.M., und Carvalho, J.L. (1979), Analysis of economic time series, New York (Academic Press).
- Newbold, P., und Davies, N. (1978), "Error mis-specification and spurious regression," IER 19, 513-519.

- Numerical Algorithms Group (1981), NAG-library manual.
- Park, R.E., und Mitchell, B.M. (1980), "Estimating the autocorrelated error model with trended data," JOE 13, 185-202.
- Park, S.B. (1982), "Some sampling properties of minimum expected loss (MEL0) estimators of structural coefficients," JOE 18, 295-311.
- Phillips, P.B. (1980), "Finite sample theory and the exact distributions of alternative estimators of the marginal propensity to consume," Review of Economic Studies 67, 183-224.
- Plasser, K. (1982), "Anwendung des IAS-SYSTEMS für den ökonomischen Modellbau und für die modellgestützte Unternehmensplanung," Angewandte Informatik 24, 541-545.
- Poirier, D.J. (1978), "The effect of the first observation in regression models with first-order autoregressive disturbances," Applied Statistics 27, 67-68.
- Prais, S.J., und Winston, D.B. (1954), "Trend estimators and serial correlation," Cowles Commission Discussion Paper No. 383, Universität Chicago.
- Quandt, R.E. (1965), "On certain small sample properties of k-Class Estimators," IER 6, 92-104.
- Rao, M.M. (1961), "Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations," Annals of Mathematical Statistics 32, 1961, 195-218.
- (1978), "Asymptotic distribution of an estimator of the boundary parameter of an unstable process," Annals of Statistics 6, 185-190.
- Rhodes, G.F., und Westbrook, M.D. (1984), "A study of estimator densities and performance under misspecification," erscheint in JOE.
- Richardson, D.H. (1968), "The exact distribution of a structural coefficient estimator," JASA 63, 1214-1226.

- , und Rohr, R.J. (1971), "Distribution of a structural t-statistic for the case of two included endogenous variables," JASA 66, 375-382.
- , und Wu, D.M. (1970), "Least squares and grouping method estimators in the errors in variables model," JASA 65, 724-748.
- , und ——— (1971), "A note on the comparison of ordinary and two-stage least squares estimates," Econometrica 39, 973-982.
- Ronning, G. (1975), "Changes of German share prices-random or not random?," Kyklos 28, 876-877.
- Rosenblatt, M. (1956), "Some regression problems in time series analysis," in: Proceedings of the third Berkeley symposium on mathematical statistics and probability, Band 1, Berkeley-Los Angeles (University of California Press), 165-186.
- Roy, R. (1977), "On the asymptotic behavior of the sample autocovariance function for an integrated moving average process," Biometrika 64, 419-421.
- , und Lefrancois, P. (1981), "Sur les moments des autocovariances échantillonnales d'un processus ARIMA," Ann. Sc. Math. Québec 5, 87-95.
- Sargan, J.D. (1975), "Gram-Charlier approximations applied to t-ratios of k-Class estimators," Econometrica 43, 169-176.
- (1976), "Econometric estimators and the Edgeworth approximation," Econometrica 44, 421-448.
- Sawa, T. (1969), "The exact finite sampling distribution of ordinary least squares and two-stage least squares estimators," JASA 64, 923-937.
- Schönfeld, P. (1971), Methoden der Ökonometrie, Band 2, München (Vahlen).
- Schmid, F. (1981), "A note on the strong consistency of least squares estimators in linear regression models," Methods of Operations Research 41, 219-221.

- Schneeweiß, H. (1971), Ökonometrie, Würzburg-Wien (Physica).
- (1976), "Consistent estimation of a regression with errors in the variables," Metrika 23, 101-115.
- (1980), "Modelle mit Fehlern in den Variablen," Methods of Operations Research 37, 41-77.
- (1982), "A simple regression model with trend and error in the exogenous variables," in: Deistler, M., Fürst, G., und Schwödiauer, G. (Hrsg.), Games, economic dynamics, and time series analysis, Wien-Würzburg (Physica).
- , und Mittag, H.-J. (1984), "Lineare Modelle mit fehlerbehafteten Daten," Würzburg-Wien (Physica).
- Seibt, H. (1968), Schätzung und Prognose bei simultanen Gleichungsmodellen, Göttingen (Vandenhoeck&Ruprecht).
- Slater, L.J. (1965), "Confluent hypergeometric functions," in: Abramowitz, M., und Stegun, J. (Hrsg.), Handbook of mathematical functions, New York (Dover), 503-536.
- Smith, V.K. (1971), "A comparative tabular survey of Monte Carlo and exact sampling studies," Australian Economic Papers 10, 196-202.
- Sonnberger, H., Havlik, P., Krämer, W., und Maurer, J. (1983), Testing for autocorrelation and model specification with the IAS-SYSTEM, Research Memorandum No. 186, Institut für Höhere Studien, Wien.
- Sperry-Univac (1973), Large Scale Systems: MATH-PACK Programmer Reference.
- Summers, R. (1965), "A capital intensive approach to the small sample properties of various simultaneous equation estimators," Econometrica 33, 1-41.
- Taylor, L.D. (1983), "A survey of finite sample results in econometrics," Econometric Reviews 2, 1-45.
- Taylor, W.E. (1981), "On the efficiency of the Cochrane-Orcutt estimator," JOE 17, 67-82.

- Theil, H. (1971), Principles of econometrics, New York (Wiley).
- Tintner, G. (1940), The variate difference method, Bloomington (Principia Press).
- (1960), Handbuch der Ökonometrie, Berlin (Springer).
- Ullah, A., und Nagar, A.L. (1974), "The exact mean of the two-stage least squares estimator of the structural parameters in an equation having three endogenous variables," Econometrica 42, 749-758.
- van der Genugten, B.B. (1982), "Asymptotic normality of least squares estimators in autoregressive linear regression models," Discussion paper No. 82-09, Subfakulteit der Econometrie, Universiteit Tilburg
- Wang, E. (1978), "Eine volkswirtschaftliche Datenbank unter FMS-8," in: Nagler-Breitenbach, I., und Schauer, H. (Hrsg.), Datenbanksysteme, Wien (Physica), 196-216.
- Waugh, F.V. (1961), "The place of least squares in econometrics," Econometrica 29, 386-396.
- White, H. (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," Econometrica 48, 817-838.
- White, J.S. (1958), "The limiting distribution of the serial correlation coefficient in the explosive case," Annals of Mathematical Statistics 29, 1188-1197.
- (1959), "The limiting distribution of the serial correlation coefficient in the explosive case II," Annals of Mathematical Statistics 30, 831-834.
- Wichern, D.W. (1973), "The behaviour of the sample autocorrelation function for an integrated moving average process," Biometrika 60, 235-239.

- Willers, R. (1978), Schwache Konsistenz von Kleinst-Quadrate Schätzern für Regressions- und Streuungsparameter in linearen Modellen, Diss. Dortmund.
- Winckler, K. (1970), "Zur Anwendung gleitender Durchschnitte auf Zeitreihen mit trendabhängiger Saisonkomponente," Jahrbücher für Nationalökonomie und Statistik 183, 533-538.
- Wold, H., und Faxer, P. (1957), "On the specification error in regression analysis," Annals of Mathematical Statistics 28, 265-267.
- Wu, C.F. (1980), "Characterizing the consistent directions of least squares," Annals of Statistics 8, 789-801.
- Yohai, V.J., und Maronna, R.A. (1977), "Asymptotic behaviour of least squares estimates for autoregressive processes with infinite variances," Annals of Statistics 5, 554-560.
- Zellner, A. (1962), "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," JASA 57, 348-368.