

Läufer, Nikolaus K. A.

Working Paper

Zu einigen Fragen bei der Anwendung der Two-stage-least-squares-Methode (2SLS)

Diskussionsbeiträge, No. 34

Provided in Cooperation with:

Department of Economics, University of Konstanz

Suggested Citation: Läufer, Nikolaus K. A. (1973) : Zu einigen Fragen bei der Anwendung der Two-stage-least-squares-Methode (2SLS), Diskussionsbeiträge, No. 34, Universität Konstanz, Fachbereich Wirtschaftswissenschaften, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/78214>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Zu einigen Fragen
bei der Anwendung der
Two-Stage-Least-Squares-
Methode (2SLS)

34

Nikolaus K.A. Läufer

April 1973

Diskussionsbeiträge des Fachbereichs Wirtschaftswissen-
schaften der Universität Konstanz

Nr. 34

I N H A L T

Seite

1.	Zur Frage der Notwendigkeit und Verfügbarkeit besonderer Schätzverfahren für simultane Gleichungssysteme	2
2.	Die Two-Stage-Least-Squares-Methode (2SLS)	6
3.	Teststatistische Fragen bei Anwendung der Two-Stage-Least-Squares-Methode (2 SLS)	12
3.1	Die Frage der Anwendbarkeit des t-Tests zur Signifikanzprüfung 2SLS-geschätzter Parameter	12
3.2	Die Frage der Anwendbarkeit des F-Tests zur Signifikanzprüfung der Gesamtregression einer 2SLS-geschätzten Strukturgleichung	15
3.3	Die Frage der Zahl der Freiheitsgrade beim t-Test	17
3.4	Notwendigkeit und Möglichkeit einer Überprüfung der Autokorrelation von Residuen	24
	Liste der Abkürzungen von Namen der zitierten Zeitschriften	32

1. Notwendigkeit und Verfügbarkeit
besonderer Schätzverfahren für
simultane Gleichungssysteme

In einem simultanen Gleichungssystem sind alle endogenen Variablen mit allen das Modell stochastisierenden Gleichungsfehlern (errors in equation bzw. shocks) korreliert.¹

Treten in einer Gleichung zwei oder mehr endogene Variable auf, dann läßt sich (nur) eine dieser Variablen willkürlich als die abhängige auswählen, und als solche auf die linke Seite einer Modellgleichung bringen. Werden die übrigen endogenen Variablen zu den erklärenden, unabhängigen Variablen auf der rechten Seite hinzugeschlagen, dann ist die Voraussetzung der Unkorreliertheit von erklärenden Variablen und zufälligem Fehlerglied (Zufallsglied) einer Gleichung, wie sie von der gewöhnlichen Kleinstquadratschätzmethode gefordert wird, nicht mehr erfüllt, weil die endogenen Variablen mit dem Zufallsfehler der Gleichung korrelieren. Wird die gewöhnliche Kleinstquadratmethode dennoch angewandt, dann erhält man solche Schätzwerte für die Parameter der Gleichung, die bei endlichem Stichprobenumfang verzerrt und bei unendlich großem Stichprobenumfang inkonsistent sind.²

¹ Das Auftreten von stochastischen Meßfehlern bewirkt, daß im allgemeinen die Korrelation zwischen dem gesamten stochastischen Glied einer Gleichung (d.h. der Summe aus Gleichungsfehler und Meßfehler von endogenen Variablen in der Gleichung) und den endogenen Variablen des Modells, insbesondere den meßfehlerfreien endogenen Variablen, geringer ist als im Fall ohne stochastische Meßfehler.

² Diese Einsicht ist zuerst von T. HAAVELMO im Zusammenhang mit der Schätzung makroökonomischer Konsumfunktionen gewonnen worden.

Die Stichprobenverteilung konsistenter Schätzwerte konzentriert sich bei zunehmendem Stichprobenumfang immer stärker um den wahren Parameterwert und bricht schließlich bei unendlich großem Stichprobenumfang über dem wahren Parameterwert zusammen.³

Für überidentifizierte Modelle⁴ gibt es mehrere Verfahren, um zu konsistenten Schätzungen zu kommen:

Vgl. T. HAAVELMO, Methods of Measuring the Marginal Propensity to Consume, in: Studies in Econometric Method, ed. W. C. HOOD und T. C. KOOPMANS, New York-London 1953, S. 75-91.

Man spricht deshalb auch vom HAAVELMO-Bias.

Vgl. St. VALAVANIS, Econometrics, An Introduction to Maximum Likelihood Methods, New York, Toronto, London 1955, S. 66.

³ Vgl. zu dieser anschaulichen Definition der Konsistenz A. S. GOLDBERGER, Econometric Theory, New York-London-Sydney 1964, S. 128.

Der Zusammenbruch besteht darin, daß die Dichtefunktion eines Parameterschätzwertes schließlich für den Wert des wahren Parameters als Argument einen unendlich großen Funktionswert, für alle anderen Argumente den Wert null annimmt.

Eine fehlerfreie Behandlung und Darstellung der Konsistenzeigenschaft findet man jedoch nicht bei GOLDBERGER sondern bei PH. J. DHRIMES, Econometrics, Statistical Foundations and Applications, New York 1970, S. 87.

⁴ Überidentifiziert sind solche Modelle, die dem für die Identifikation notwendigen und hinreichenden Rangkriterium genügen und bei denen mindestens eine Gleichung des Modells ferner dem notwendigen, aber nicht hinreichenden Abzählkriterium sogar in der strengeren, echten Ungleichungsform ($>$) und nicht nur in der schwächeren, unechten Ungleichungsform (\geq) genügt.

Zuerst zum Rangkriterium: Sei A die Koeffizientenmatrix eines linearen simultanen Gleichungssystems. A enthalte also die Koeffizienten sowohl der endogenen als auch der vorherbestimmten Variablen des Modells. Sei A' jene Matrix, die aus der Matrix A durch Streichen gewisser Spalten hervorgeht. Gestrichen werden solche Spalten, die in einer

bestimmten Zeile Koeffizienten haben, welche nicht a priori gleich Null sind. Die maßgebende Zeile ist jene, in der die Koeffizienten derjenigen Gleichung des Modells stehen, deren Identifizierbarkeit gerade untersucht werden soll. Diese Gleichung ist genau dann identifizierbar, wenn der Rang von A' gleich $n-1$ ist. n ist dabei die Zahl der endogenen Variablen (Gleichungen) des Modells.

Das als Identifikationsbedingung notwendige, nicht aber hinreichende Abzählkriterium verlangt, daß die Zahl der aus einer einzelnen Gleichung ausgeschlossenen, aber im Modell überhaupt vorkommenden vorherbestimmten Variablen größer oder gleich (\geq) sein muß der um eins verminderten Zahl der in der Gleichung vorkommenden endogenen Variablen. Wird neben dem Rangkriterium das Abzählkriterium in der Gleichungsform von allen Gleichungen des Modells erfüllt, dann ist das (ganze) Modell exakt identifiziert, und Schätzwerte für die Parameter des Modells, d.h. seine Struktur, sind aus den geschätzten Koeffizienten der reduzierten Form des Modells durch rationale Operationen eindeutig ableitbar. Ganz analog unterscheidet man anhand des Abzählkriteriums zwischen exakt identifizierten und überidentifizierten Gleichungen eines (nicht unbedingt vollständig formulierten) Modells. Ist das Modell unvollständig formuliert (spezifiziert), dann setzt man dabei stillschweigend voraus, daß eine dem Abzählkriterium genügende Gleichung auch dem Rangkriterium genügt. Ist umgekehrt das Abzählkriterium bei einer Gleichung nicht erfüllt, dann ist bei ihr notwendigerweise auch das Rangkriterium nicht erfüllt, und deshalb ist diese Gleichung, und gleichzeitig das ganze zugehörige Modell, nicht identifiziert bzw. identifizierbar. Rang- und Abzählkriterium haben wir hier durchweg auf Gleichungen bzw. Modelle bezogen, für die, außer bei definitorischen Gleichungen und Gleichgewichtsbedingungen, lediglich variablenausschließende a-priori-Restriktionen verwendet werden. Vgl. zum Rang- und Abzählkriterium irgendein Standardwerk der Ökonometrie wie z.B. J. JOHNSTON, *Econometric Methods*, New York 1963, S. 240ff., bes. anschaulich auch St. VALAVANIS, a.a.O., S. 85ff.

Die Full-Information-Maximum-Likelihood-Methode (FI), die linearisierte Maximum-Likelihood-Methode (LML) und die Three-Stage-Least-Squares-Methode (3SLS), alle drei als Systemschätzmethoden, ferner die Limited-Information-Maximum-Likelihood (LI)- und die Two-Stage-Least-Squares- (2SLS)-Methode als Einzelgleichungsschätzmethoden.⁵ Bei den Einzelgleichungsschätzmethoden können die Parameter einer einzelnen, überidentifizierten Gleichung auch dann geschätzt werden, wenn die übrigen Gleichungen, d.h. der restliche Teil des Modells, nicht spezifiziert worden sind. Es ist lediglich notwendig,

⁵ Die LI- und 2SLS-Methoden sind Spezialfälle der allgemeineren Schätzmethode der k-Klasse (k-class estimator) von H. THEIL. Vgl. dse., Economic Forecasts and Policy, 2. Aufl., Amsterdam 1961, S. 231 f. Bei dieser Schätzmethode kommt man stets dann zu konsistenten Schätzungen, wenn der Wahrscheinlichkeitslimes von $(k - 1)$ gleich null ist, d.h. wenn $\text{plim}(k - 1) = 0$. Bei LI ist das der Fall, bei 2SLS, wegen $k = 1$, ebenfalls. Die OLS gehören ebenfalls zur k-Klasse, ihr k-Wert ist jedoch von vornherein gleich null, so daß $\text{plim}(k - 1) = -1$, d.h. ungleich null, OLS also inkonsistent ist. Die Schätzmethode der k-Klasse ist selbst wiederum ein Spezialfall der allgemeineren Methode der doppelten k-Klasse von A. L. NAGAR. Siehe dessen Beitrag: Double k-Class Estimators of Parameters in Simultaneous Equations and Their Small Sample Properties, IER 3/1962. Eine nützliche tabellarische Übersicht über ausgewählte, bekannte mathematische Eigenschaften der genannten Schätzmethoden, insbesondere über ihre Konsistenzeigenschaften, findet man bei C. F. CHRIST, Econometric Models and Methods, New York, London, Sidney 1966, S. 466. In diesem Werk findet man auch eine pädagogisch meisterhafte Darstellung der verschiedenen Schätzmethoden und ihrer gegenseitigen Beziehungen.

daß man einen (für die Zwecke der Identifikation genügenden) Teil der vorherbestimmten Variablen aus dem übrigen Teil des Modells kennt.⁶

2. Die Two-Stage-Least-Squares-Methode (2SLS)

Wir werden uns im folgenden näher mit der Two-Stage-Least-Squares-Methode (2SLS) befassen, weil sie die noch immer nicht infragegestellten Vorzüge einer Einzelgleichungsschätzmethode für Simultanmodelle⁷ mit rechentechnischer Einfachheit verbindet.⁸

⁶ Die Frage der Identifizierbarkeit einer einzelnen Gleichung aus einem im übrigen unbekanntem simultanen System wird dadurch leicht zum Objekt von bloßen Annahmen.

⁷ Im Gegensatz zu Rekursiv-Modellen nach WOLD.

⁸ Während in der Monte-Carlo-Studie von R. M. SUMMERS (1965) die 2SLS-Methode als relativ beste Methode unter FI, LI und OLS abschnitt, insbesondere wenn man Rechenaufwand, Spezifikationsfehler und Multikollinearität der vorherbestimmten Variablen mit in die abwägende Entscheidung einbezog, schrumpften in einer Untersuchung von R. E. QUANDT (1965) die Vorteile von 2SLS gegenüber OLS auf eine höhere Konzentration der Verteilung der 2SLS-Schätzungen innerhalb des Bereichs von $\pm 20\%$ des wahren Parameterwertes zusammen, während falsche Vorzeichen und hohe Fehler bei 2SLS relativ häufiger als bei OLS vorkamen, so daß die Verteilung der 2SLS-Schätzwerte an den Enden stärker und die Streuung deshalb insgesamt größer war als bei OLS.
Vgl. R. M. SUMMERS, A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators, Em 33/1965 und R. E. QUANDT, On Certain Small Sample Properties of k-Class Estimators, IER 6/1965. Siehe ferner

Ferner enthält sie als Spezialfall die gewöhnliche Kleinstquadratmethode (OLS) und ist deshalb für Vergleiche zwischen 2SLS und OLS, vom rechentechnischen Standpunkt aus, sehr geeignet. Außerdem gestattet 2SLS in der Frage der Signifikanz der Schätzwerte eine, wenn auch nur eingeschränkt berechnete, Anlehnung an die Tests der gewöhnlichen Regressionsanalyse, welches wiederum günstige rechentechnische Konsequenzen mit sich bringt (OLS läßt sich auch hier als Spezialfall von 2SLS betrachten). Wir werden die 2SLS-Methode auch dann anwenden, wenn die zu schätzende Gleichung exakt identifiziert ist. Das geschieht deshalb, weil die in diesem Fall ebenfalls mögliche Methode der indirekten kleinsten Quadrate (ILS) zwar zu eindeutigen Schätzwerten für die Parameter der exakt identifizierten Gleichung führt, ihre Anwendung aber die noch zu erörternden Testmöglichkeiten für die Parameter der Strukturgleichung (im Gegensatz zu den Parametern der reduzierten Form) ausschließt.

Beim 2SLS-Verfahren werden die Parameter b_i, c_i der Gleichung (1)

$$y_1 = b_2 y_2 + b_3 y_3 + \dots + b_g y_g + c_1 z_1 + \dots + c_k z_k + u$$

y_i = eingeschlossene endogene Variable $i = 1, \dots, g$

z_j = eingeschlossene vorherbestimmte Variable
 $j = 1, \dots, k$

die zusammenfassende Wiedergabe bei C. F. CHRIST, a.a.O., S. 464ff., insbes. S. 477-481.

Im Vergleich zu LI sind bei 2SLS weniger Matrizen zu invertieren und iterative Rechenprozesse fallen bei 2SLS weg. Veränderungen des simultanen Modells machen bei Einzelgleichungsmethoden nicht notwendigerweise die Neuschätzung des ganzen Modells erforderlich wie bei den Systemmethoden FI, 3SLS usw.

theoretisch in zwei Schritten geschätzt.⁹

Im ersten Schritt werden die OLS-Schätzwerte \hat{y}_i der endogenen unter den erklärenden rechtsstehenden Variablen, d.h. für y_i $i = 2, \dots, g$ berechnet, wobei alle vorherbestimmten Variablen des Modells, nicht nur die in der zu schätzenden Gleichung (1) eingeschlossenen, als Regressoren verwendet werden:

$$\hat{y}_i = y_i - \hat{v}_i = \quad (2)$$

$$\hat{a}_{i1}z_1 + \dots + \hat{a}_{ik}z_k + \hat{a}_{i,k+1}z_{k+1} + \dots + \hat{a}_{il}z_l$$

$$i = 2, \dots, g$$

z_{k+h} = aus der zu schätzenden Gleichung ausgeschlossene vorherbestimmte Variable des Modells

$$h = 1, \dots, (l - k)$$

\hat{a}_{ij} = OLS-geschätzte Parameter

⁹ Für u nehmen wir an, daß alle Voraussetzungen erfüllt sind, die man üblicherweise verlangt. Üblicherweise wird verlangt, daß der Erwartungswert eines Zufallsgliedes gleich null ist, daß Zufallsglieder verschiedener Zeitpunkte unkorreliert sind, unabhängig davon, ob sie zur selben oder zu verschiedenen Gleichungen gehören. "Gegenwärtige" Korrelationen zwischen Zufallsgliedern verschiedener Gleichungen dürfen von null verschieden sein. Die "gegenwärtigen" Kovarianzen und Varianzen müssen jedoch zeitunabhängig, d.h. zeitlich konstant sein (Homoskedastizität). Für ein ganzes simultanes Gleichungssystem wird ferner gefordert, daß die gegenwärtigen Zufallsglieder mit allen vorherbestimmten Variablen unkorreliert sind. Vgl. C. F. CHRIST, *Econometric Models and Methods*, New York, London, Sidney 1966, S. 433. Wir haben im folgenden bis zur Formel (5) alle Zeitindizes der Variablen weggelassen.

Im zweiten Schritt werden die OLS-Schätzwerte \hat{y}_i zusammen mit den eingeschlossenen vorherbestimmten Variablen als Regressoren zur Schätzung der Parameter b_i und c_i der Gleichung (1) wieder nach der OLS-Methode verwendet.

Die in zweiter Stufe mit OLS zu schätzende Gleichung lautet also:

$$y_1 = b_2 \hat{y}_2 + b_3 \hat{y}_3 + \dots \quad (3)$$

$$+ b_g \hat{y}_g + c_1 z_1 + \dots + c_k z_k + w$$

worin

$$w = u + b_2 \hat{v}_2 + b_3 \hat{v}_3 + \dots + b_g \hat{v}_g \quad (4)$$

und w als stochastisches Glied mit allen anderen rechtsstehenden Variablen (\hat{y}_i und z_j) in (3) unkorreliert¹⁰ ist, so daß die OLS-Methode von den Voraussetzungen her tatsächlich anwendbar ist.

Auch läßt sich die bereits oben behauptete Konsistenz der Schätzungen nachweisen.¹¹ Die 2SLS-Schätzungen¹² sind ferner asymptotisch normal verteilt, wenn alle

¹⁰ u ist voraussetzungsgemäß unkorreliert mit allen vorherbestimmten Variablen, infolgedessen auch mit einer beliebigen Linearkombination derselben, und deshalb auch mit den Variablen \hat{y}_i , $i = 2, \dots, g$. Die \hat{v}_i ($i = 2, \dots, g$) sind ebenfalls unkorreliert mit allen vorherbestimmten Variablen, also auch unkorreliert mit den Linearkombinationen \hat{y}_i derselben, sowie mit den eingeschlossenen vorherbestimmten Variablen. w ist als Linearkombination der Variablen u und \hat{v}_i ($i = 2, \dots, g$), welche mit den \hat{y}_i und den eingeschlossenen z_j nicht korrelieren, selbst unkorreliert mit den \hat{y}_i und z_j .

¹¹ Vgl. C. F. CHRIST, a.a.O., S. 438-440.

¹² Zur 2SLS-Methode generell siehe C. F. CHRIST, a.a.O., S. 432ff.

vorherbestimmten Variablen exogen sind.¹³ Für den Fall, in dem unter den vorherbestimmten Variablen auch verzögert endogene vorkommen, wird die asymptotische Normalität bis jetzt lediglich vermutet, sie wurde noch nicht exakt bewiesen.¹⁴

Praktisch werden die \hat{y}_i ($i = 2, \dots, g$) nicht explizit ausgerechnet, sondern die beiden Stufen werden zu einer einzigen Schätzfunktion verbunden, die in ihrer Grundstruktur der OLS-Schätzfunktion äußerlich gleicht.

Die OLS-Schätzfunktion, auf die Gleichung (3) angewendet, lautet:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = \begin{pmatrix} Y_2' \hat{Y}_2 & Y_2' Z_1 \\ Z_1' Y_2 & Z_1' Z_1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{Y}_2' \\ Z_1' \end{pmatrix} y_1 \quad (5)$$

¹³ Wenn wir für die Strukturgleichungen normalverteilte Zufallsglieder unterstellen, sind die Schätzungen sogar asymptotisch effizient, sofern man nur solche Schätzmethoden vergleicht, die dieselbe (unvollständige) a-priori-Information verwenden wie 2SLS, und sofern unter den vorherbestimmten Variablen nur exogene Variable vorkommen. Vgl. C. F. CHRIST, a.a.O., S. 466.

¹⁴ Vgl. C. F. CHRIST, a.a.O., S. 441 und S. 466.

Darin bedeutet:

b = Spaltenvektor der Koeffizienten der "erklärenden" endogenen Variablen in der Gleichung (1)
[($g - 1$) Komponenten]

c = Spaltenvektor der Koeffizienten der vorherbestimmten Variablen in der Gleichung (1)
(k Komponenten)

\hat{b}, \hat{c} = entsprechende Schätzvektoren

y_1 = Beobachtungsspaltenvektor der zu erklärenden endogenen Variablen (N Beobachtungen)

Z_1 = Beobachtungsmatrix der eingeschlossenen vorherbestimmten Variablen (N Zeilen, k Spalten)

Y_2 = Beobachtungsmatrix der rechtsstehenden, "erklärenden" endogenen Variablen (N Zeilen, ($g - 1$) Spalten)

\hat{Y}_2 = Matrix der (erststufigen) Schätzwerte für die Beobachtungen der "erklärenden" endogenen Variablen, geschätzt unter Verwendung aller im System enthaltenen vorherbestimmten Variablen; oder einfach: Schätzmatrix für Y_2 (N Zeilen, ($g - 1$) Spalten)

Ersetzt man nun in der Schätzfunktion (5) die Matrix \hat{Y}_2 durch die ihr entsprechende erststufige OLS-Schätzfunktion, d.h. setzt man

$$\hat{Y}_2 = Z\hat{A}' = Z(Z'Z)^{-1}Z'Y_2 \quad (6)$$

worin Z die Beobachtungsmatrix (mit N Zeilen und 1 Spalten) aller vorherbestimmten Variablen des Systems und \hat{A} die Matrix (mit ($g - 1$) Zeilen und 1 Spalten) der \hat{a}_{ij} in (2) darstellen, in die Formel (5) ein,

dann erhält man die Formel (7), in der auf der rechten Seite nur noch Beobachtungswerte der Variablen vorkommen. (7)

$$\begin{pmatrix} \hat{b} \\ \hat{c} \end{pmatrix} = \begin{pmatrix} Y_2'Z(Z'Z)^{-1}Z'Y_2 & Y_2'Z_1 \\ Z_1'Y_2 & Z_1'Z_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_2'Z(Z'Z)^{-1}Z' \\ Z_1' \end{pmatrix} Y_1$$

Diese Formel der 2SLS-Schätzfunktion besitzt dieselbe Grundstruktur wie die Formel der konventionellen OLS-Schätzfunktion.

3. Teststatistische Fragen bei Anwendung der Two-Stage-Least-Squares-Methode (2SLS)

3.1 Die Frage der Anwendbarkeit des t-Tests zur Signifikanzprüfung 2SLS-geschätzter Parameter

Für die 2SLS-Schätzwerte \hat{b}_i und \hat{c}_j der Parameter b_i und c_j der Gleichung (1) lassen sich approximative Varianzen und Kovarianzen schätzen.¹⁵

Bezeichnen wir diese Schätzungen mit $\hat{\sigma}^2(\hat{b}_i)$ und $\hat{\sigma}^2(\hat{c}_j)$, dann sind die folgenden Größen, wegen

¹⁵ Dazu benutzt man die Formeln für die nachweisbar existierenden asymptotischen Varianzen und Kovarianzen der 2SLS-Schätzungen und ersetzt alle Wahrscheinlichkeitsgrenzwerte durch ihre entsprechenden Stichprobenwerte. Zum Nachweis der Existenz der asymptotischen Varianz-Kovarianz-Matrix vgl. C. F. CHRIST, a.a.O., S. 441ff. Zur Schätzung der approximativen Varianz-Kovarianz-Matrix vgl. dse. S. 444.

der Konsistenz und asymptotischen Normalität der 2SLS-Schätzungen \hat{b}_i und \hat{c}_j , approximativ normalverteilt¹⁶ mit Mittelwert 0:

$$\frac{\hat{b}_i - b_i}{\hat{\sigma}(\hat{b}_i)}, \quad \frac{\hat{c}_j - c_j}{\hat{\sigma}(\hat{c}_j)} \quad (8)$$

$$i = 2, \dots, g \quad j = 1, \dots, k$$

Die Schätzung erfolgt (siehe auch weiter unten den Abschnitt über Autokorrelation), mit einer Ausnahme, wie beim OLS-Verfahren, wenn man dieses auf die Gleichung (3) anwendet, was ja bei 2SLS theoretisch tatsächlich geschieht. Das bedeutet: in der OLS-Varianz-Kovarianz-Formel

$$\hat{\sigma}^2(X'X)^{-1}$$

werden die Spalten der Matrix X für die Zwecke der 2SLS mit den "Beobachtungs"-Vektoren der Variablen $\hat{y}_2, \hat{y}_3, \dots, z_k$ auf der rechten Seite von (3) identifiziert, d.h. die Matrix X läßt sich auch als $X = (\hat{Y}_2 Z_1)$ schreiben mit der gleichen Bedeutung der Symbole \hat{Y}_2 und Z_1 wie in Formel (5). Als Schätzung von $\hat{\sigma}^2$ wird aber nun nicht $\hat{\sigma}_w^2$ nach (3) berechnet, sondern $\hat{\sigma}_u^2$ verwendet, wobei \hat{u}_t anhand von (1), d.h. mittels y_2, y_3, \dots und nicht mittels $\hat{y}_2, \hat{y}_3, \dots$ berechnet wird. Zur Berechnung von \hat{u}_t werden also die beobachteten y_i -, nicht die erststufigen \hat{y}_i -Werte verwendet.

¹⁶ Die folgenden Bemerkungen zur approximativen Normalität entnehmen wir größtenteils C.F. CHRIST, a.a.O., S. 515.

Approximativ und nicht exakt normalverteilt mit Mittelwert null sind diese Größen aus mehreren Gründen. Erstens sind die b_i und c_j nur approximativ, und nicht exakt normalverteilt. Zweitens sind die Erwartungswerte \hat{b}_i und \hat{c}_j möglicherweise verzerrt und die Erwartungswerte der Größen (8) sind deshalb nicht notwendigerweise gleich null. Drittens sind die $\hat{\sigma}^2(\)$ nur Schätzwerte der approximativen, nicht der exakten Standardabweichungen der Schätzungen \hat{b}_i und \hat{c}_j . Ferner sind die $\hat{\sigma}^2(\)$ vermutlich nur approximativ χ^2 -verteilt. Deshalb sind die Größen (8) nicht exakt, sondern nur approximativ t-verteilt und deshalb, da die t-Verteilung selbst eine Approximation der Normalverteilung darstellt bzw. umgekehrt, auch nur approximativ normalverteilt.

Infolge dieser Überlegungen halten wir die Anwendung des t-Tests, so wie er üblicherweise zur Überprüfung von OLS-geschätzten Parametern auf signifikante Unterschiede gegenüber dem hypothetischen Wert 0 oder einem anderen vorgegebenen Wert verwendet wird, in approximativer Bedeutung auch bei 2SLS für berechtigt und sinnvoll. Denn wenn man, wie z.B. CHRIST, bereits von approximativer Normalverteilung der Größen (8) spricht, dann sichern wir uns noch zusätzlich ab, indem wir bei der approximativen Prüfung von Signifikanzen auch noch die beschränkte Zahl der Freiheitsgrade unserer Schätzungen beachten und die Signifikanzprüfung anhand der t-Verteilung statt der Normalverteilung vornehmen. Denn bekanntlich entspricht die Normalverteilung der t-Verteilung genau dann, wenn die Zahl der Freiheitsgrade unendlich groß ist.

Und bei beschränkter, endlicher Zahl der Freiheitsgrade besitzt die t-Verteilung eine größere Varianz als die Normalverteilung, so daß, im Vergleich zum Normalverteilungstest, beim t-Test eine größere Abweichung des geschätzten Parameters von der Nullhypothese erforderlich ist, damit diese Abweichung bei

gegebener Irrtumswahrscheinlichkeit als signifikant gelten kann.

3.2 Die Frage der Anwendbarkeit des F-Tests zur Signifikanzprüfung der Gesamtregression einer 2SLS-geschätzten Strukturgleichung

Die Frage, ob der bei OLS übliche F-Test zur Überprüfung eines multiplen Korrelationskoeffizienten (R) übertragbar ist auf 2SLS zur Überprüfung eines R nach der Definition¹⁷

$$R^2 = 1 - \frac{\sum_t \hat{u}_t^2}{\sum_t (y_{1t} - \bar{y}_1)^2} \quad (9)$$

ist zu verneinen. Diese Antwort gründet sich weniger auf die Beobachtung¹⁸, daß der F-Test, im Gegensatz zum t-Test, sehr empfindlich auf Abweichungen von der Normalverteilungsvoraussetzung der zugrundeliegenden Variablen¹⁹ reagiert, als vielmehr auf die Beobachtung,

¹⁷ Darin werden bei 2SLS die \hat{u}_t anhand der Gleichung (1) und nicht aufgrund der Gleichung (3) geschätzt, sollten also nicht mit \hat{v}_t verwechselt werden. Ansonsten ist R hier analog zum multiplen Korrelationskoeffizienten und R^2 analog zum multiplen Bestimmtheitskoeffizienten von OLS definiert.

¹⁸ Vgl. J. PFANZAGL, Allgemeine Methodenlehre der Statistik, Bd. II, Berlin 1962, S.190.

¹⁹ Unter den "zugrundeliegenden" Variablen verstehen wir hier die stochastischen Glieder u_t .

daß bei 2SLS die Größe R^2 gemäß der Definition (9) negativ und R deshalb imaginär werden kann²⁰. Während bei OLS stets die Gleichung²¹

$$\sum_{t=1}^N y_{1t}^2 = \sum_{t=1}^N \hat{y}_{1t}^2 + \sum_{t=1}^N \hat{u}_t^2 \quad (10)$$

gilt, ist sie bei 2SLS im allgemeinen nicht mehr erfüllt. Statt dessen gilt bei 2SLS allgemein mit seltenen Ausnahmen

$$\sum_t y_{1t}^2 \neq \sum_t \hat{y}_{1t}^2 + \sum_t \hat{u}_t^2 \quad (11)$$

Ohne weiteres ist bei 2SLS deshalb auch die Ungleichung

$$\sum_t \hat{u}_t^2 > \sum_t y_{1t}^2 \quad (12)$$

und damit, aus (9) folgend, ein negativer R^2 -Wert möglich²². Die Größe

$$F = \frac{R^2/(k-1)}{(1-R^2)/(N-k)} \quad (13)$$

²⁰ Diese "Befürchtung" äußert z.B. auch R.L. BASSMANN in seinem: Letter to the Editor, Em 30/1962.

²¹ Wir verwenden bei y_1 in den Ausdrücken (10), (11) und (12), aber nicht im Ausdruck (9) Abweichungen vom Mittelwert.

²² Im übrigen sind nach meinem "Empfinden" die Wahrscheinlichkeiten des Auftretens negativer R^2 -Werte je nach der Schätzmethode verschieden und bei LI größer als bei 2SLS.

worin k die Zahl der zu schätzenden Parameter einer Gleichung²³ und N den Stichprobenumfang (Zahl der Beobachtungen) darstellt, kann deshalb bei 2SLS ohne weiteres, d.h. mit nicht verschwindender Wahrscheinlichkeit, negativ werden, was den Annahmen einer F -Verteilung widerspricht und die Anwendbarkeit des F -Tests ausschließt.

3.3 Die Frage der Zahl der Freiheitsgrade beim t -Test

Bei der approximativen Anwendung des t -Tests auf 2SLS-geschätzte Parameter taucht die Frage nach der korrekten Anzahl der Freiheitsgrade auf. Für diese Zahl gibt es zwei verschiedene Berechnungsformen.

Form 1:

- a) Zahl der Beobachtungen (N) minus Zahl der in der Gleichung rechtsstehenden, "erklärenden" Variablen ($g - 1 + k$), d.h. $N - (g - 1 + k) = N - g - k + 1$.
- b) Zahl der Beobachtungen (N) minus der um 1 verminderten Zahl der in der Gleichung enthaltenen endogenen (g) und enthaltenen vorherbestimmten (k) Variablen, d.h. $N - [(g + k) - 1] = N - g - k + 1$.

²³ In der Zahl k ist auch das Absolutglied der (linearen) Gleichung enthalten, d.h. der Parameter jener Erklärungsvariablen, die über die ganze Beobachtungsperiode unverändert den Wert 1 annimmt und deshalb tatsächlich nichts zur Erklärung der Varianz der zu erklärenden oder abhängigen Variablen beiträgt.

Form 2:

Zahl der Beobachtungen (N) minus Zahl aller vorherbestimmten Variablen des Modells bzw. des Gleichungssystems (2), d.h.: $N - 1$.

Für exakt identifizierte Gleichungen, also solchen, bei denen die Anzahl der im System (Modell), nicht jedoch in der (betrachteten) Modellgleichung enthaltenen vorherbestimmten Variablen ($l - k$), gerade gleich der Zahl der endogenen Variablen der betrachteten Gleichung minus 1 ist, d.h. gleich der Zahl ($g - 1$) der rechtsstehenden, "erklärenden" endogenen Variablen der Gleichung, wenn also $(l - k) = (g - 1)$ gilt, dann fallen die beiden Formen zusammen.²⁴ Es kann hier deshalb keine Meinungsverschiedenheit geben. Bei überidentifizierten Gleichungen, also solchen, bei denen die Zahl der ausgeschlossenen vorherbestimmten Variablen ($l - k$) größer ist als die Zahl der eingeschlossenen endogenen minus 1, wenn also $(l - k) > (g - 1)$, dann gibt es jedoch abweichende Auffassungen. So halten sich J. JOHNSTON²⁵ und C. F. CHRIST²⁶ an Form 1, und darin folgen wir ihnen, R. L. BASSMANN²⁷ dagegen hält sich an Form 2.

²⁴ Aus $(l - k) = (g - 1)$ folgt $l = k + g - 1$
und $(N - 1) = N - (k + g - 1) = N - k - g + 1$.

²⁵ a.a.O., S. 272

²⁶ a.a.O., S. 518f. (dort implizit), siehe jedoch auch S. 515 wegen der Bemerkung: "The appropriate number of degrees of freedom ... is not clear". (Auslassung durch uns), so daß die in der Praxis vorherrschende Form 1 auch von CHRIST nicht als einwandfrei betrachtet wird.

²⁷ Vgl. R. L. BASSMANN, The Computation of Generalized Classical Estimates of Coefficients in a Structural Equation, Em 27/1959, insbes. S. 77.

Zur Ermittlung der Zahl der Freiheitsgrade kommt es im folgenden darauf an, die verschiedenen (unabhängigen) Beschränkungen zu ermitteln, denen die Residuen \hat{u}_t im Rahmen des 2SLS-Verfahrens unterworfen sind. Das geschieht indirekt auf dem Weg über die Beschränkungen, denen die Residuen der ersten Stufe von 2SLS, \hat{v}_t , und die Residuen der zweiten Stufe von 2SLS, \hat{w}_t , unterworfen sind. Zwischen den Residuen der beiden Verfahrensstufen besteht die aus (4) folgende Beziehung (14),

$$\hat{u}_t = \hat{w}_t - \hat{b}_2 \hat{v}_{2t} - \hat{b}_3 \hat{v}_{3t} - \dots - \hat{b}_g \hat{v}_{gt} \quad (14)$$

worin \hat{u}_t die Schätzung des Zufallsgliedes der betrachteten, 2SLS-geschätzten Gleichung eines Systems darstellt.

Aus den Eigenschaften der gewöhnlichen Kleinstquadratmethode (OLS) und aus dem Umstand, daß die erste Stufe des 2SLS-Verfahrens nichts anderes ist, als eine Anwendung von OLS, leiten wir für die aus der ersten Stufe des 2SLS-Verfahrens hervorgehenden \hat{v}_{it} ab, daß sie mit allen vorherbestimmten Variablen unkorreliert sind.²⁸

²⁸ Die OLS-Schätzfunktion lautet in der üblichen Symbolik: $\hat{\beta} = (X'X)^{-1}X'y$,

worin X die Beobachtungsmatrix der erklärenden, y den Beobachtungsspaltenvektor der zu erklärenden Variablen und $\hat{\beta}$ den Spaltenvektor der Parameterschätzwerte darstellen.

Wenden wir nun sinngemäß die Symbole von (2) an, dann lauten die OLS-Schätzfunktionen für $a_i = (a_{i1}, \dots, a_{i1})$, d.h. für den Vektor der Koeffizienten a_{ij} ($j = 1, \dots, 1$) der Erklärungsgleichung für die i-te endogene Variable (erste Stufe von 2SLS):

$$\hat{a}'_i = (Z'Z)^{-1}Z'y_i \quad i = 2, \dots, g,$$

\hat{w}_t , welches aus der zweiten Stufe von 2SLS hervor-
 geht, ist mit \hat{y}_{it} ($i = 2, \dots, g$) und mit
 z_{jt} ($j = 1, \dots, k$), also mit einem Teil der vorher-
 bestimmten Variablen des Systems, unkorreliert. Das
 folgt ebenfalls aus den Eigenschaften von OLS und aus
 dem Umstand, daß die zweite Stufe von 2SLS ebenfalls
 eine Anwendung von OLS darstellt. Folglich sind die
 vorherbestimmten Variablen zumindest teilweise, näm-
 lich die z_{jt} ($j = 1, \dots, k$), mit \hat{w}_t unkorreliert.
 Wenn nun \hat{w}_t auch mit den z_{jt} ($j = k + 1, \dots, l$),
 also den ausgeschlossenen, vorherbestimmten Variablen
 unkorreliert wäre, dann wären sowohl \hat{w}_t als auch die
 \hat{v}_{it} und daraus nach (14) folgend \hat{u}_t mit allen vorher-
 bestimmten Variablen unkorreliert und Form 2 wäre
 korrekt. Leider läßt sich für \hat{w}_t lediglich feststellen,
 daß es außer mit den eingeschlossenen vorherbestimmten
 Variablen z_{jt} ($j = 1, \dots, k$) nur noch mit den
 y_{it} ($i = 2, \dots, g$) d.h. mit $(g - 1)$ linear unabhän-
 gigen Linearkombinationen aller vorherbestimmten
 Variablen unkorreliert ist. Wegen $(l-k) > (g - 1)$,

worin Z die Beobachtungsmatrix aller vorherbestimmten
 Variablen und y_i den Beobachtungsspaltenvektor der
 i -ten endogenen Variablen darstellt. Der Spaltenvektor
 der Residuen \hat{v}_i ist definiert als $\hat{v}_i = y_i - Z\hat{a}_i$.
 Daraus folgt für das innere Produkt des Vektors \hat{v}_i
 mit den Beobachtungsvektoren Z_j der vorherbestimm-
 ten Variablen z_{jt} ($j = 1, \dots, l$), wobei die Z_j die
 Spalten der obigen Matrix Z darstellen, also
 $Z = (Z_1 \dots Z_l)$ gilt:

$$\begin{aligned}
 Z'\hat{v}_i &= Z'y_i - Z'Z\hat{a}_i \\
 &= Z'y_i - (Z'Z)(Z'Z)^{-1}Z'y_i \\
 &= Z'y_i - Z'y_i = 0 \quad i = 2, \dots, g,
 \end{aligned}$$

d.h. die \hat{v}_i sind orthogonal zu den Z_j bzw. die \hat{v}_{it}
 korrelieren nicht mit den z_{jt} .

d.h. weil die Zahl der ausgeschlossenen vorherbestimmten Variablen (voraussetzungsgemäß) größer ist als die Zahl der rechtsstehenden, erklärenden endogenen Variablen, d.h. größer als die Zahl der fraglichen Linearkombinationen, können vorherbestimmte Variable existieren, mit denen \hat{w}_t und deshalb auch \hat{u}_t tatsächlich korreliert.²⁹ Auf jeden Fall wird bei

²⁹Zur strengeren Fundierung dieser Feststellung bilden wir folgendes Gleichungssystem:

$$\hat{w}'\hat{y}_i = \hat{w}'Z\hat{a}_i \quad i = 2, \dots, g.$$

\hat{w} ist hier als Spaltenvektor mit N Komponenten aufzufassen ($t=1, \dots, N$).

\hat{y}_i ist die i-te Spalte der $N \times (g-1)$ -Matrix \hat{Y}_2 mit der Bedeutung wie in (5).

\hat{a}_i sind linear unabhängige Zeilenvektoren mit 1-Komponenten der $(g-1) \times 1$ -Matrix \hat{A} in (6).

$Z = (Z_1, Z_2)$ ist die $N \times 1$ -Beobachtungsmatrix aller vorherbestimmten Variablen.

Z_1 die $N \times k$ -Matrix der eingeschlossenen, Z_2 die $N \times (1-k)$ -Matrix der ausgeschlossenen vorherbestimmten Variablen.

Wir teilen den Zeilenvektor \hat{a}_i der Koeffizienten der erststufigen Regression analog zu Z auf:

$$\hat{a}_i = (\hat{a}_i^1, \hat{a}_i^2).$$

Wir wissen von der zweiten Stufe des 2SLS-Verfahrens, daß $\hat{w}'Z_1 = 0$ und $\hat{w}'\hat{y}_i = 0$ für $i = 2, \dots, g$.

Daraus folgt aus dem obigen Gleichungssystem

$$\hat{w}'\hat{y}_i = \hat{w}'(Z_1, Z_2) \begin{pmatrix} \hat{a}_i^1 \\ \hat{a}_i^2 \end{pmatrix} = 0 + \hat{w}'Z_2\hat{a}_i^2 = 0$$

$$i = 2, \dots, g$$

Das Gleichungssystem $\hat{w}'Z_2\hat{a}_i^2 = 0 \quad i = 2, \dots, g$

ist homogen und besteht aus $(g-1)$ Gleichungen in $(1-k)$ "unbekannten" Kovarianzen zwischen \hat{w}_t und den ausgeschlossenen vorherbestimmten Variablen. Die $(g-1)$ linear unabhängigen Zeilenvektoren \hat{a}_i^2 enthalten die (konstanten) Koeffizienten des linearen Gleichungssystems. Ist nun $(1-k) > (g-1)$, dann ist die Zahl der "unbekannten" Kovarianzen größer als die Zahl der Gleichungen. Es gibt dann nicht-triviale Lösungen des Gleichungssystems, d.h. die

überidentifizierten Gleichungen eine solche von null verschiedene Korrelation nicht durch irgendwelche Restriktionen des 2SLS-Verfahrens verhindert. Die Zahl der Restriktionen, denen die \hat{u}_t unterworfen sind, ist bei überidentifizierten Gleichungen kleiner als 1. Folglich ist die Zahl der Freiheitsgrade größer als $(N - 1)$. Das spricht gegen die Form 2.

Wie bereits gesagt, ist \hat{u}_t entsprechend dem Inhalt der zweiten Stufe von 2SLS unkorreliert mit k (der l) vorherbestimmten Variablen und mit $(g - 1)$ Linearkombinationen der l vorherbestimmten Variablen. Für die \hat{w}_t bestehen keine weiteren Restriktionen. Dasselbe gilt für \hat{u}_t , obwohl die \hat{v}_{it} , also ein Teil der (additiven) Komponenten von \hat{u}_t , nicht nur $k + g - 1$, sondern l Restriktionen unterworfen sind. Das bedeutet, daß die Zahl der Freiheitsgrade nicht unter $N - (g - 1) - k$ liegt.

Sie kann aber auch nicht darüber liegen. Denn das würde voraussetzen, daß sich unter den $(g - 1)$ (linear unabhängigen) Linearkombinationen aus den l vorherbestimmten Variablen wenigstens eine befände, die mit einer der k eingeschlossenen vorherbestimmten Variablen vollständig (d.h. mit Wert 1) korreliert wäre. Und das wiederum würde den Fall vollständiger Multikollinearität bedeuten und schon die erste Stufe des 2SLS-Verfahrens zusammenbrechen lassen.

Andererseits gibt es im Fall einer überidentifizierten Gleichung, bei der ja $(l - k) > (g - 1)$ gilt, genügend Spielraum dafür, daß aus den l vorherbestimmten Vari-

"unbekannten" Kovarianzen sind nicht notwendigerweise gleich null.

Ist aber $(l - k) = (g - 1)$, dann gibt es notwendigerweise nur eine Lösung, die zugleich trivial ist, d.h. alle "unbekannten" Kovarianzen sind dann gleich null.

ablen $(g - 1)$ Linearkombinationen gebildet werden können, die von den k eingeschlossenen vorherbestimmten Variablen (linear) unabhängig sind, so daß das 2SLS-Verfahren nicht aus Gründen vollständiger Multikollinearität notwendigerweise zusammenbrechen müßte.

Infolgedessen gibt Form 1 (auch) im Falle überidentifizierter Gleichungen die korrekte Zahl der Freiheitsgrade an. Verwendet man im Fall überidentifizierter Gleichungen die geringere Zahl der Freiheitsgrade nach Form 2, dann läßt man bei der approximativen Anwendung der Signifikanzteste eine zusätzliche Vorsicht walten, - die Nullhypothese für die Parameterwerte wird weniger leicht zurückgewiesen. Die Wahrscheinlichkeit eines Fehlers erster Art sinkt dadurch.

Das von mir programmierte Rechenprogramm für das 2SLS-Verfahren macht von Form 1 Gebrauch, vermeidet also die in Form 2 bei überidentifizierten Gleichungen enthaltene zusätzliche Vorsicht.

Bei exakt identifizierten Gleichungen entspricht die Form 1 der Form 2 und beide Formen entsprechen dann der wahren Zahl der Freiheitsgrade. Bei überidentifizierten Gleichungen gibt allein Form 1 die richtige Zahl der Freiheitsgrade an.³⁰

³⁰ Die Umrechnung der approximativen t -Größen ist einfach. Bezeichnen wir mit t_i den t -Wert (des t -Tests) mit f_i Freiheitsgraden nach Form i , $i = 1, 2$, dann gilt nach Definition der t -Werte (siehe [8]) die Beziehung:

$$t_1 = t_2 \cdot \frac{f_1}{f_2}$$

Ferner gilt:

1. $f_2 < f_1$ genau dann, wenn $t_2 < t_1$, und $t_2 < t_1$ genau dann, wenn $(1 - k) > (g - 1)$.

2. $f_2 = f_1$ genau dann, wenn $t_2 = t_1$, und $t_2 = t_1$ genau dann, wenn $(1 - k) = (g - 1)$.

3.4 Notwendigkeit und Möglichkeit einer Überprüfung der Autokorrelation von Residuen

Bei autokorrelierten Zufallsgliedern hat die Anwendung von OLS auf Einzelgleichungssysteme mehrere unangenehme Konsequenzen. Erstens gelten die OLS-Schätzformeln für die Varianz der Parameterschätzungen nicht. Zweitens wird wahrscheinlich die in diese Formel eingehende Varianz des Zufallsgliedes systematisch unterschätzt, und drittens sind die (wahren) Varianzen der Parameterschätzungen nicht mehr minimal, d.h. eine Optimumeigenschaft von OLS geht dann verloren.³¹

Unterschätzung der Streuung der Parameterschätzungen führt bei Anwendung des t- und F-Tests leicht vor- schnell zu einem signifikanten Resultat, d.h. zu Verwerfungen der Nullhypothese, wo sie bei der gewählten Irrtumswahrscheinlichkeit tatsächlich noch aufrechterhalten werden müßte. Bevor man bei OLS die Nullhypothese endgültig verwirft, sollte man also prüfen, ob die Residuen als Schätzwerte der unbeobachteten Zufallseinflüsse autokorreliert sind. Wenn ja, dann ist das Testergebnis, zusätzlich zum leicht kalkulierbaren und in Irrtumswahrscheinlichkeiten ausgedrückten Testrisiko, mit nicht bzw. nur umständlich kalkulierbarer Unsicherheit behaftet.

³¹ Vgl. J. JOHNSTON, a.a.O., S. 187ff. Außerdem besteht bei OLS und Eingleichungssystemen die Gefahr, daß die Autokorrelation sich mit anderen Hindernissen für die OLS-Methode paart und dann zu nicht voraussehbaren Konsequenzen für die Schätzergebnisse führt (z.B. Umkehrung des Vorzeichens eines infolge verzögerter Variabler (allein) zu erwartenden negativen Bias in einen positiven). Vgl. dse., a.a.O., S. 215f.

Mit denselben unangenehmen Konsequenzen für die Anwendung des t-Tests ist eine Unterschätzung der Streuung der Parameterschätzwerte infolge autokorrelierter Zufallsglieder wahrscheinlich auch bei 2SLS zu erwarten. Und wie bei OLS gibt es dafür zwei Gründe. Erstens ist die übliche Formel für die asymptotischen Varianzen und Kovarianzen der Parameterschätzwerte bei autokorrelierten Störgliedern der betrachteten Strukturgleichung eines Mehrgleichungssystems nicht mehr korrekt, folglich auch nicht mehr die approximative Schätzung der Varianz der Parameterschätzwerte, die sich auf diese asymptotische Formel stützt. Zweitens wird auch hier wahrscheinlich die Varianz des Zufallsgliedes systematisch unterschätzt.

Autokorrelierte Zufallsglieder behindern jedoch nicht nur die approximative Anwendbarkeit des t-Tests, sondern wirken sich auch nachteilig auf andere Eigenschaften der 2SLS-Schätzfunktion aus. In Modellen mit verzögert endogenen unter den vorherbestimmten Variablen können die 2SLS-Schätzungen ihre Konsistenz-eigenschaft verlieren, wenn die Zufallsglieder verschiedener Perioden nicht unkorreliert sind. Aber selbst dann, wenn alle vorherbestimmten Variablen exogene Variablen sind, kann bei autokorrelierten Zufallsgliedern die Konsistenzeigenschaft verlorengehen, sicherlich aber geht die Eigenschaft der asymptotischen Effizienz der 2SLS-Schätzungen, falls vorhanden³², verloren.³³

³² Die asymptotische Effizienz von 2SLS ist dann gegeben, wenn die Menge der vorherbestimmten Variablen nur exogene Variablen enthält und wenn die Zufallsglieder außerdem gemeinsam normalverteilt sind.

Ein für Einzelgleichungssysteme geeigneter Test zur Überprüfung der Nullhypothese, daß die unbeobachteten Zufallsglieder (nicht zu verwechseln mit ihren Schätzwerten, den Residuen) nicht autokorreliert sind, wurde von DURBIN und WATSON³⁴ entwickelt. Dieser Test stützt sich auf die Größe d , die definiert wird als

$$d = \frac{\sum_{t=2}^N (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^N (\hat{u}_t - \bar{\hat{u}})^2} \quad (15)$$

worin \hat{u}_t die Residuen, oder allgemeiner, Schätzwerte der Zufallsvariablen u_t darstellen. Handelt es sich um OLS- oder 2SLS-Residuen, dann gilt³⁵ $\bar{\hat{u}} = 0$ und d lautet dann

$$d = \frac{\sum_{t=2}^N (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^N (\hat{u}_t)^2} \quad (16)$$

Vgl. wegen der übrigen Voraussetzungen die Beschreibung des von C. F. CHRIST sogenannten "model 3 (normal)" in C. F. CHRIST, a.a.O., S. 350ff., insbes. S. 356, sowie auf S. 466 die Tabelle über die Eigenschaften von 2SLS unter den Voraussetzungen des "model 3 (normal)".

³³ Siehe C. F. CHRIST, a.a.O., S. 481.

³⁴ Siehe J. DURBIN und G. S. WATSON, Testing for Serial Correlation in Least Squares Regression I, Bm 37/1950, ferner dieselben, Testing ... Regression II, Bm 38/1951.

³⁵ Das folgt für 2SLS aus Gleichung (14) und aus der Gültigkeit von $\bar{\hat{w}} = \bar{\hat{v}}_i = 0$ bei den OLS-Stufen von 2SLS.

DURBIN und WATSONs Test für d überprüft die Autokorrelation der Zufallsglieder u_t anhand ihrer Schätzwerte \hat{u}_t . Obige Nullhypothese bezieht sich also auf u_t und nicht auf \hat{u}_t , obwohl sie anhand von d , d.h. letztlich mittels \hat{u}_t getestet wird. Darin unterscheidet sich dieser Test vom Test mit dem ursprünglichen VON-NEUMANN-Quotienten³⁶, welcher definiert wird als

$$\frac{\delta^2}{S^2} = \frac{\sum_{t=2}^N (u_t - u_{t-1})^2 / (N - 1)}{\sum_{t=1}^N (u_t - \bar{u}_t)^2 / N} \quad (17)$$

HART und v. NEUMANN³⁷ haben die Verteilung dieser Größe vertafelt unter der Annahme, daß u_t normalverteilt ist. Der VON-NEUMANN-Quotient untersucht die Autokorrelation an den unbeobachteten Zufallsgliedern selbst, nicht aber an den entsprechenden Schätzwerten. Anders formuliert:

Der ursprüngliche VON-NEUMANN-Quotient kann, auf die Residuen angewandt, d.h. \hat{u}_t an Stelle von u_t in die Formel für $\frac{\delta^2}{S^2}$ eingesetzt, prinzipiell nicht zuverlässig die Autokorrelation der unbeobachteten Zufallsglieder selbst, d.h. der u_t , überprüfen, sondern nur eben die der Residuen.

Monte-Carlo-Studien von ORCUTT und COCHRANE³⁸ haben diesen Mangel eindeutig gezeigt.

³⁶ Siehe John von NEUMANN, Distribution of the Ratio of the Mean Square Successive Difference to the Variance, AMS 12/1941, sowie dse., A Further Remark Concerning the Distribution of the Ratio of the Mean Square Successive Difference to the Variance, AMS 13/1942.

³⁷ Siehe B. I. HART und J.v. NEUMANN, Tabulation of the Probabilities for the Ratio of the Mean Squares Successive Difference to the Variance, AMS 13/1942.

In zu vielen Fällen führte die Anwendung des VON-NEUMANN-Quotienten auf die Residuen (\hat{u}) zur Annahme der Nullhypothese fehlender Autokorrelation, der in der Studie vorgegebenen und damit bekannten Zufallsglieder (u), wo sie in Wahrheit, - so die Versuchsanordnung -, hätte abgelehnt werden müssen. Die Verteilung des VON-NEUMANN-Quotienten ist also nicht geeignet zur Entdeckung der Autokorrelation von unbeobachteten Zufallsgliedern anhand von Schätzwerten für dieselben. Es läßt sich mit ihr entweder die Autokorrelation der Residuen anhand einer Stichprobe von Residuen oder aber die Autokorrelation der unbeobachteten Zufallsglieder anhand einer Stichprobe von unbeobachteten Zufallsgliedern (d.h. gar nicht) überprüfen.

Leider herrscht in der Literatur eine gewisse Verwirrung, dadurch daß die DURBIN-WATSON-Statistik d auch als VON-NEUMANN-Quotient bezeichnet bzw. der VON-NEUMANN-Quotient mit Hilfe der DURBIN-WATSON- d -Formel definiert wird.³⁹

Der d -Test von DURBIN und WATSON wurde ausgearbeitet für Regressionsgleichungen (Einzelgleichungssysteme), in denen alle erklärenden Variablen exogene Variablen sind. Er gilt deshalb streng genommen nicht für Gleichungen, in denen verzögerte Werte der abhängigen Variablen als erklärende Variablen vorkommen, und ferner nicht durchgängig für einzelne Gleichungen eines simultanen Gleichungssystems.

³⁸ Siehe G. H. ORCUTT und D. COCHRANE, A Sampling Study of the Merits of Autoregressive and Reduced Form Transformations in Regression Analysis, JASA 44/1949

³⁹ Siehe z.B. H. THEIL und A.L. NAGAR, Testing the Independence of Regression Disturbances, JASA 56/1961, insbes. die Fn. 2 auf S. 793.
Eine klare Unterscheidung findet man dagegen bei C.F. CHRIST, a.a.O., S. 523ff., auf die wir uns größtenteils stützen.

Die Verteilung der Statistik d hat eine Spannweite von 4 und ⁴⁰⁾ hängt zunächst ab von der Zahl der Beobachtungen (N) und der Zahl der erklärenden Variablen außer dem Absolutglied ($k - 1$). Da sie ferner abhängig ist von den Beobachtungswerten für y , der abhängigen Variablen, und für z_i ($i = 1, \dots, k$) den (exogenen) erklärenden Variablen, ist es nicht gelungen, sie allgemeingültig anzugeben. Die Verteilung von d wird statt dessen eingegrenzt von den Verteilungen für zwei andere statistische Größen, nämlich d_u und d_l , die nicht von den Beobachtungswerten selbst, wohl aber (auch) von der Zahl der Beobachtungen und der Zahl der erklärenden Variablen abhängen.

Um die Alternativhypothese positiver Autokorrelation gegen die Nullhypothese keiner Autokorrelation zu prüfen, stellt man deshalb für eine ausgewählte Irrtumswahrscheinlichkeit in Abhängigkeit von N und ($k - 1$) die zugehörigen kritischen Werte für d_u und d_l fest. Diese Werte grenzen das der gewählten Irrtumswahrscheinlichkeit entsprechende kritische d ein.⁴¹ Ist dann das berechnete d größer als oder gleich d_u , dann nimmt man die Nullhypothese an, ist d kleiner als oder gleich d_l , dann wird die Nullhypothese zugunsten der Alternativhypothese zurückgewiesen. Wenn die Konstellation $d_l < d < d_u$ auftritt, ist das Testergebnis nicht schlüssig.

Um die Hypothese negativer Autokorrelation zu testen, verwendet man aus Symmetriegründen die Größen $(4 - d_u)$ und $(4 - d_l)$. Zu einer ausgewählten Irrtumswahrscheinlichkeit bilden also die zugehörigen Werte d_l und d_u eine untere und obere Schranke für

⁴⁰ Die Verteilung schrumpft mit zunehmendem Stichprobenumfang ($N \rightarrow \infty$) auf den Wert 2 zurück.

⁴¹ Unter dem kritischen d verstehen wir die sonst so genannte Grenze des Annahmebereichs der Nullhypothese, im englischen auch acceptance level genannt.

das kritische d beim Test positiver Autokorrelation. Entsprechend bilden $(4 - d_u)$ und $(4 - d_l)$ die untere und obere Schranke für das kritische d beim Test negativer Autokorrelation. Es existieren Tabellen⁴² für einseitige Fragestellungen, die wegen der Symmetrie der Verhältnisse auch für zweiseitige Tests anwendbar sind.

Orientiert man sich in der Praxis allein an den Schranken d_u und $4 - d_u$, dann erhöht man etwas die Irrtumswahrscheinlichkeit, d.h. man wird etwas häufiger Autokorrelation annehmen, wenn tatsächlich keine gegeben ist.⁴³ Beim Test auf Autokorrelation des Zufallsgliedes in Einzelgleichungen mit der verzögert abhängigen unter den erklärenden Variablen wird jedoch die ausschließliche Anwendung von d_u (bzw. $4 - d_u$) geradezu empfohlen.⁴⁴

DURBIN's Untersuchung⁴⁵ der Frage der Anwendbarkeit des d -Tests auf einzelne Gleichungen simultaner Gleichungssysteme hat ergeben, daß die Anwendung der

⁴² Siehe J. DURBIN und G. S. WATSON, Testing ... Regression II, a.a.O.

⁴³ Es gibt jedoch Anhaltspunkte dafür, daß das kritische d näher bei d_u als bei d_l liegt, siehe C. F. CHRIST, a.a.O., S. 527.

⁴⁴ Siehe C. F. CHRIST, a.a.O., S. 527f. und unsere vorhergehende Fußnote.

⁴⁵ Siehe J. Durbin, Testing for Serial Correlation in Systems of Simultaneous Regression Equation Bm 44/1957.

DURBIN-WATSON-Statistik d auf strukturelle Residuen und die Anwendung der Tabellen für d_1 und d_u korrekt ist, sofern die Menge der vorherbestimmten Variablen nur exogene Variablen enthält und sofern die Strukturgleichung gerade oder exakt identifiziert ist. Bei überidentifizierten Strukturgleichungen, und wiederum unter der Voraussetzung, daß keine verzögert endogenen Variablen im Modell vorkommen, gilt der d -Test approximativ. Ein für diesen Fall genauer Test wird zusätzlich von DURBIN angegeben.

Läßt man verzögert endogene Variable zu, dann bleibt die näherungsweise Gültigkeit erhalten. Leider nimmt DURBIN bei seinen Überlegungen an, daß die Gleichungen mit der Limited-Information-Methode geschätzt werden. Wir glauben aber, daß wir keine großen Fehler machen, wenn wir DURBINs Resultate (ohne den genauen Test für überidentifizierte Gleichungen) auch auf die 2SLS-Methode übertragen.⁴⁶ In jedem Fall bringen wir unseren Resultaten bei der approximativen Signifikanzprüfung der 2SLS-Schätzungen eine angemessene Skepsis entgegen, wenn wir uns an d_u allein statt an d_u und d_1 orientieren.

⁴⁶ Bei der Handhabung der Tabelle für d_1 und d_u gilt hier als Zahl der erklärenden Variablen die ^uGesamtzahl der im Gleichungssystem insgesamt enthaltenen vorherbestimmten Variablen ohne die Dummy-Variable für das Absolutglied der betrachteten Gleichung. Siehe J. DURBIN, a.a.O., S. 377

Liste der Abkürzungen von Namen der zitierten
Zeitschriften

AMS	Annals of Mathematical Statistics
Bm	Biometrika
Em	Econometrica
IER	International Economic Review
JASA	Journal of the American Statistical Association