

Dynamische Modelle: Schätzen und Modellwahl

Kapitel 18

Angewandte Ökonometrie / Ökonometrie III
Michael Hauser

Inhalt

- ▶ Dynamische Modelle und autokorrelierte Fehler
- ▶ Tests auf Autokorrelation in den Residuen
- ▶ Behebung von Autokorrelation in den Residuen
- ▶ Spezifikation von Modellen, Modellwahl

Dynamische Modelle und autokorrelierte Fehler

Schätzung eines AR(1): korrekt spezifiziert

Das autoregressive Modell erster Ordnung lautet, $|\alpha_1| < 1$

$$Y_t = \alpha + \alpha_1 Y_{t-1} + u_t, \quad u_t \dots \text{WN}$$

Das Mittel von Y ergibt sich als langfristiges GGW ($\bar{Y} = Y_t$).

$$\mu_y = \alpha / (1 - \alpha_1)$$

Der OLS Schätzer für α_1 ist

$$\hat{\alpha}_1 = \hat{\rho}_1(Y)$$

der Stichprobenautokorrelationskoeffizient 1.Ord von Y .

Er ist *in small samples verzerrt*, aber *konsistent* und *asy normal verteilt*.

Das gilt auch in ADL(r, s), Fehlerkorrektur-, und verwandte Modelle.

$u_t \dots \text{WN}$ ist gleichbedeutend mit $u_t \text{ iid}(0, \sigma_u^2)$.

Schätzung eines AR(1): misspezifiziert

Ist die Störung nicht weißes Rauschen, so sind die OLS Schätzer auch *asymptotisch verzerrt*.

Bsp: (vgl. Quasidifferenzen unten)

$$Y_t = \alpha_1 Y_{t-1} + u_t, \quad u_t = \rho_1 u_{t-1} + v_t, \quad v_t \dots \text{WN}$$

Wir multiplizieren das Modell für Y_{t-1} mit ρ_1

$$\rho_1 Y_{t-1} = \alpha_1 \rho_1 Y_{t-2} + \rho_1 u_{t-1} \quad \text{und subtrahieren von } Y_t.$$

$$Y_t - \rho_1 Y_{t-1} = \alpha_1 Y_{t-1} - \alpha_1 \rho_1 Y_{t-2} + u_t - \rho_1 u_{t-1}$$

mit $u_t - \rho_1 u_{t-1} = v_t$ WN, erhalten wir

$$Y_t = (\alpha_1 + \rho_1) Y_{t-1} - \alpha_1 \rho_1 Y_{t-2} + v_t$$

Das ist ein AR(2) Modell!

Als AR(1) Modell (Ann. u WN) geschätzt erhalten wir für den Koeffizienten vor Y_{t-1} jedenfalls nicht $(\alpha_1 + \rho_1)$, da es misspezifiziert ist.

Quasidifferenzen

Das Subtraktionsergebnis läßt sich auch wie folgt anschreiben:

$$(Y_t - \rho_1 Y_{t-1}) = \alpha_1 (Y_{t-1} - \rho_1 Y_{t-2}) + (u_t - \rho_1 u_{t-1})$$

in Quasidifferenzen

$$Y_t^* = \alpha_1 Y_{t-1}^* + v_t, \quad v_t \dots \text{WN}$$

mit

$$Y_t^* = Y_t - \rho_1 Y_{t-1}, \quad X_t^* = X_t - \rho_1 X_{t-1}, \quad u_t^* = u_t - \rho_1 u_{t-1}$$

Die Variablen X^* , Y^* und u^* heißen **Quasidifferenzen**.

Das Modell $Y_t = \beta_0 + \beta_1 X_t + u_t$ in Quasidifferenzen lautet:

$$Y_t^* = \beta_0^* + \beta_1 X_t^* + v_t$$

β_0^* wird analog berechnet: $\beta_0^* = \beta_0(1 - \rho_1)$.

Quasidifferenzen beheben die Autokorrelation 1.Ordnung in den Fehlern.

Tests auf autokorrelierte Residuen in dynamischen Modellen

Autokorrelierte Residuen

Ursachen für autokorrelierte Residuen sind *Fehlspezifikationen* des Modells:

- ▶ fehlende erklärende Variable (wie im statischen Modell), oder falsche Wahl der Lagordnung,
- ▶ im Speziellen falsche Wahl der Lagordnungen der verzögerten endogenen Variablen,
- ▶ y -Variable ist nicht stationär, z.B. $I(1)$ oder linearer Trend, x -Variable sind aber stationär.

Tests auf autokorrelierte Residuen

Der Durbin-Watson Test ist nur bei deterministischen Regressoren anwendbar, *nicht* jedoch bei stochastischen bzw. verzögerten endogenen Variablen!

Tests auf autokorrelierte Residuen bei stochastischen Regressoren sind:

- ▶ Durbins's h
- ▶ Breusch-Godfrey-Test

Wir diskutieren die Tests am Bsp. des ADL(1,0):

$$Y_t = \alpha_1 Y_{t-1} + \beta_0 X_t + u_t$$

mit $u_t = \rho_1 u_{t-1} + v_t$, $v_t \dots$ WN.

Durbin's h , Test auf Autokorrelation 1. Ordnung

$$H_0: \rho_1 = 0$$

$$H_1: \rho_1 \neq 0$$

Teststatistik:

$$h = (1 - 0.5d) \sqrt{n/[1 - n\text{Var}(\hat{\alpha}_1)]} \stackrel{a}{\sim} N(0, 1)$$

d ist die Durbin-Watson Statistik. Für sie gilt ($\hat{\rho}_1 = r_1$):

$$d \approx 2(1 - r_1)$$

Somit ist

$$h \approx r_1 \sqrt{n/[1 - n\text{Var}(\hat{\alpha}_1)]}$$

Bem: Diese Statistik kann man nicht berechnen, wenn $\text{Var}(\hat{\alpha}_1) > 1/n$.

Breusch-Godfrey-Test auf Autokorr 1.Ordnung

$$H_0: \rho_1 = 0$$

$$H_1: \rho_1 \neq 0$$

Testablauf:

1. OLS Regression des Modells, z.B. ADL(1,0). Sie erhalten die Residuen e_t .
2. Regression e_t auf Y_{t-1}, X_t, e_{t-1} . Erhalten R_e^2 .
3. $LM = nR_e^2$

Teststatistik:

$$LM \stackrel{a}{\sim} \chi^2(1)$$

Dies ist nur ein Test auf Autokorrelation der Ordnung 1. Das Korrelogramm zeigt möglicherweise Autokorrelation höherer Ordnung in den Residuen.

*Verfahren zur Behebung von autokorrelierten Residuen:
Cochran-Orcutt*

Cochran-Orcutt Verfahren

Das Verfahren von Cochran-Orcutt dient zur Aufnahme der Autokorrelationsstruktur der Residuen in das Modell.

Angenommen einer der Tests zeigt Autokorrelation in den Residuen an. Wir gehen für das ADL(1,0) Modell $Y_t = \alpha_1 Y_{t-1} + \beta_0 X_t + u_t$ Schritt-weise vor:

1. Wir bilden die Quasidifferenzen mit $r_1 = \hat{\rho}_1$ für alle Variable des Modells.

$$Y_t^* = Y_t - r_1 Y_{t-1} \quad X_t^* = X_t - r_1 X_{t-1}$$

2. Wir schätzen das Modell nun in Quasidifferenzen anstelle der ursprünglichen Variablen.

$$Y_t^* = \alpha_1 Y_{t-1}^* + \beta_0 X_t^* + u_t^*$$

3. Wir testen die neuen Fehler, u_t^* , auf Autokorrelation.

Cochran-Orcutt Verfahren, Alternative

- ▶ Sind die neuen Fehler u_t^* unkorreliert. Ende der Prozedur.
- ▶ Sind sie autokorreliert, wenden wir die Prozedur nochmals auf das zuletzt geschätzte Modell an.

Alternativ kann z.B. das ADL(1,0) Modell mit autokorrelierten Fehlern in EViews mit einem Befehl geschätzt werden.

$$y \ y(-1) \ x \ AR(1)$$

Hier wird die Cochran-Orcutt Prozedure iterativ (mehrmals) durchgeführt. Der AR(1) Koeffizient gibt den Schätzer für ρ_1 .

Instrumenten Variablen, IV, Schätzung

Wird eine IV Schätzung um Autokorrelation in den Residuen korrigiert, so sind in der Instrumentenliste alle verzögerten Variablen (die für die explizite Berechnung der Quasidifferenzen gebraucht werden würden) ebenfalls anzugeben.

Andernfalls ist der Schätzer nicht konsistent.

Modellspezifikation, Modellwahl: from-general-to-specific

Modellwahl

Wir behandeln hier:

- ▶ überspezifizierte Modelle
Was ist die Konsequenz, wenn falsche Variable im Modell verwendet werden?
- ▶ unterspezifizierte Modelle
Was ist die Konsequenz, wenn Variable fehlen?
- ▶ from-general-to-specific
Welche Variable sollen in ein Regressionsmodell aufgenommen werden, wenn mehrere zur Auswahl stehen?

Modellwahl: Überspezifizierte Modelle

Angenommen das *wahre* Modell ist

$$Y_t = a + bX_t + u_t$$

aber sie schätzen das **überspezifizierte** Modell

$$Y_t = a + bX_t + cZ_t + v_t$$

Die Variable Z kann eine dritte Variable sein, ein verzögertes Y oder ein verzögertes X .

Im Durchschnitt wird der geschätzte Parameter c null sein, und die Parameter a und b *unverzerrt*. Bei Durchführung einer Prognose muß allerdings mit einer erhöhten Prognosevarianz gerechnet werden, da die zusätzliche Variable Z_t noise erzeugt. ($\hat{c} \neq 0$ i.A.)

Modellwahl: Unterspezifizierte Modelle

Angenommen das *wahre* Modell ist

$$Y_t = a + bX_t + cZ_t + u_t$$

aber sie schätzen das **unterspezifizierte** Modell

$$Y_t = a + bX_t + v_t$$

Die relevante Variable Z fehlt (z.B. die Variable ist nicht in der Theorie vorgesehen, wurde vergessen, oder die Daten sind nicht verfügbar). Die Schätzer von a und b werden versuchen den fehlenden Effekt von Z aufzunehmen. (Es sei denn Z ist orthogonal zu X und hat eine Mittel 0.)

Daher werden die Schätzer wie auch die Prognosen **verzerrt** sein.

Modellwahl: complete search

Ang. sie haben eine Vielzahl von Variablen (z.B. proxies) zur Verfügung, bzw. sie wissen nicht, welche Lags sie in ein dynamisches Modell aufnehmen sollen.

Variante 1, complete search: Sie probieren alle möglichen Kombinationen durch.

- ▶ Wenn sie wissen, dass das wahre Modell eines der untersuchten ist, und sie wenden das SBC an, dann finden sie das korrekte Modell mit $n \rightarrow \infty$. Für endliche n wählt man mal ein zu kleines, mal ein zu großes Modell, oder schließt mal eine falsche Variable mit ein. Aber wenn der Stichprobenumfang groß wird, treffen sie das wahre Modell.
- ▶ Wenden sie das AIC an, so tendieren sie dazu in kleinen Stichproben und wenn $n \rightarrow \infty$ ein zu großes Modell zu wählen, mit den oben skizzierten Vor- und Nachteilen.

Modellwahl: from-general-to-specific

Variante 2, from-general-to-specific:

Können sie nicht alle Möglichkeiten durchprobieren, so starten sie mit dem größten Modell und reduzieren sukzessive, in dem sie die 'schlechteste' Variable in der nächsten Modellvariante ausschließen. Sie führen dies solange durch bis sie nach dem jeweiligen Informationskriterium kein besseres Modell finden. Dieses Verfahren hat die gleichen Optimalitätseigenschaften wie das complete search.

Angenommen ein geschätztes Modell lautet

$$Y_t = \hat{a} + \hat{b}Y_{t-1} + \hat{c}X_t + \hat{d}Z_t + \hat{u}_t, \quad AIC = 7.444$$

Sie wollen wissen, welche der Variablen im nächsten Schritt eliminiert werden soll. Nun schätzen sie alle 3 Modell-Varianten mit nur 2 Variablen auf der rechten Seite.

from-general-to-specific (Fs.)

Regressoren	Y_{t-1}, X_t	Y_{t-1}, Z_t	X_t, Z_t
AIC	7.100	7.200	7.500

Das erste Modell hat das kleinste AIC. Dieses ist kleiner als das oben gefundene 7.444. Daher ist das vorläufig beste Modell ($Y_t | Y_{t-1}, X_t$).
Sie schließen daher Variable Z im nächsten Schritt aus.

Nun probieren wir alle 1-Variablen Varianten von $\{Y_{t-1}, X_t\}$

Regressoren	Y_{t-1}	X
AIC	7.150	7.250

Alle AIC Werte sind größer als die des zuletzt gefundenen. Daher finden wir als 'bestes' Modell ($Y_t | Y_{t-1}, X_t$).

Exkurs: AR(1) versus I(1)

Stationärer autoregressiver Prozess 1.Ordnung, AR(1)

Für einen autoregressiver Prozess der 1.Ordnung, AR(1), y_t , gibt es verschiedene Darstellungen:

- ▶ Rekursive Darstellung, als Regression (stoch. Differenzgleichung)
 $y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t, \quad u_t \text{ WN}, \quad |\alpha_1| < 1.$
- ▶ Rekursive Darstellung, Zeitreihennotation (Mittelwert-korrigierte Reihe)
 $(y_t - \mu) = \alpha_1 (y_{t-1} - \mu) + u_t \quad u_t \text{ WN}, \quad |\alpha_1| < 1. \quad E(y_t) = \mu.$

Die Beziehung zwischen α_0 dem Interzept und dem Erwartungswert μ ist

$$\mu = \frac{\alpha_0}{1 - \alpha_1}$$

Wenn Sie in EViews z.B. ein ARMA(1,1)-Modell schätzen wird diese Darstellung verwendet. Die Konstante im Output ist das geschätzte Mittel.

$y \subset \text{AR}(1) \text{ MA}(1)$

Stationärer autoregressiver Prozess 1.Ordnung, AR(1)

- ▶ Rekursive Darstellung mit Lagoperator ($Ly_t = y_{t-1}$)

$$(1 - \alpha_1 L)y_t = \alpha_0 + u_t$$

Der Prozess ist stationär, wenn die Nullstelle (Wurzel, root) des charakteristischen Polynoms $(1 - \alpha_1 z) = 0$, i.e. $z = 1/\alpha_1$, außerhalb des Einheitskreises liegt. D.h.: $|z| > 1$, bzw. $|\alpha_1| < 1$.

Stationärer autoregressiver Prozess 1.Ordnung, AR(1)

- ▶ Explizite Darstellung mit y_0 im Startzeitpunkt $t = 0$ und $\alpha_0 = \mu = 0$

$$y_t = \alpha_1^t y_0 + \sum_{i=0}^{t-1} \alpha_1^i u_{t-i}$$

- ▶ Explizite Darstellung mit y_0 in $t = 0$ und $\alpha_0 \neq 0$ bzw. $\mu \neq 0$

$$y_t = \alpha_0 \sum_{i=0}^{t-1} \alpha_1^i + \alpha_1^t y_0 + \sum_{i=0}^{t-1} \alpha_1^i u_{t-i}$$

- ▶ Explizite Darstellung für den Startzeitpunkt gegen $(-\infty)$ und $|\alpha_1| < 1$.

$$y_t = \frac{\alpha_0}{1 - \alpha_1} + \sum_{i=0}^{\infty} \alpha_1^i u_{t-i}$$

Für $|\alpha_1| < 1$ verschwindet der Effekt des Anfangswertes y_0 mit $t \rightarrow \infty$, was gleichbedeutend mit der Verschiebung des Anfangszeitpunktes gegen $(-\infty)$ ist.

Mit $|\alpha_1| < 1$ existiert $\lim_{t \rightarrow \infty} \sum_{i=0}^{t-1} \alpha_1^i = 1/(1 - \alpha_1)$.

Stationärer autoregressiver Prozess 1.Ordnung, AR(1)

Der (unbedingte) Erwartungswert und die (unbedingte) Varianz eines AR(1)-Prozesses sind

$$E y_t = \frac{\alpha_0}{1 - \alpha_1} = \mu, \quad V y_t = E(y_t - \mu)^2 = \frac{1}{1 - \alpha_1^2} \sigma_u^2 = \gamma(0)$$

Beide existieren und sind konstant.

Die Autokovarianzfunktion, ACF, ist für $k = 0, 1, 2, 3, \dots$

$$\gamma(k) = E(y_t - \mu)(y_{t-k} - \mu)$$

Sie fällt geometrisch für $|\alpha_1| < 1$ gegen Null.

Das Korrelogramm ist die Folge von Autokorrelationskoeffizienten

$$\rho(k) = \gamma(k)/\gamma(0) = \alpha_1^k$$

Random Walk, integrierter Prozess der Ordnung 1, I(1)

- ▶ Der Random Walk (RW), y_t , mit Drift c , rekursiv angeschrieben ist

$$y_t = c + y_{t-1} + u_t \quad u_t \dots \text{WN}$$

Ist $c = 0$, ist es ein RW ohne Drift.

- ▶ In der expliziten Darstellung ausgehend von Startzeitpunkt $t = 0$, $y_0 = y_0$

$$y_t = y_0 + c \cdot t + \sum_{i=1}^t u_i$$

Die bedingte Erwartung und die bedingte Varianz ausgehend von $t = 0$ sind

$$E(y_t|y_0) = y_0 + c \cdot t \quad V(y_t|y_0) = \sigma_u^2 \cdot t$$

Der Driftparameter c verursacht einen linearen Trend in der bedingten Erwartung. Das Essentielle hier ist aber die Summe der iid u 's, die eine linear steigende Varianz verursachen.

Verschiebt man den Anfangszeitpunkt gegen $(-\infty)$, so geht die bedingte Varianz gegen ∞ .

Die Stichprobenautokorrelation (SACF) fällt langsam gegen Null.

Test auf einen integrierten Prozess der Ordnung 1, Dickey-Fuller Test

Wir untersuchen, ob der Koeffizient $\phi = 1$ oder $\phi < 1$ ist. (Einseitiger Test)

$$y_t = c + \phi y_{t-1} + u_t$$

Wir ziehen auf beiden Seiten y_{t-1} ab und schätzen mit OLS.
(Δy_t ist unter beiden Hypothesen H_0 und H_A stationär.)

$$\Delta y_t = c + (\phi - 1)y_{t-1} + u_t$$

$$H_0 : \phi = 1 \quad H_A : \phi < 1$$

oder

$$H_0 : \phi - 1 = 0 \quad H_A : \phi - 1 < 0$$

Die Dickey-Fuller-Teststatistik (DF) ist die t -Statistik des geschätzten Koeffizienten $(\widehat{\phi - 1})$. Es wird so wie beim t -test auf Null getestet.

Test auf einen integrierten Prozess der Ordnung 1, Dickey-Fuller Test

Die DF-Statistik ist aber nicht-standard verteilt und daher tabelliert.

Die Verteilung unterscheidet sich je nachdem,

- ▶ ob kein Interzept im Modell auftritt

$$y_t = \phi y_{t-1} + u_t$$

- ▶ ob ein Interzept vorhanden ist

$$y_t = c + \phi y_{t-1} + u_t$$

- ▶ oder ob $I(1)$ gegen einen linearen Trend getestet werden soll.

$$y_t = c + \phi y_{t-1} + \beta t + u_t$$