

Prognose

Kapitel 5

Ökonometrie I
Michael Hauser

Inhalt

- ▶ Prognose, allgemein
- ▶ Prognosebeispiel
- ▶ Punktprognose
- ▶ Prognosefehler
- ▶ Intervallprognose
- ▶ Mean square error, Prognosegüte

Prognose

Die Prognose (griechisch, prognosis – wörtlich 'das Vorwissen', die 'Voraus-Kenntnis'), Vorhersage oder Voraussage ist eine Aussage über Ereignisse, Zustände oder Entwicklung in der 'Zukunft'.

Neben der Beschreibung von Zusammenhängen, Testen von Hypothesen, ist die Prognose die dritte wichtige Aufgabe von Modellen. Auch ist die Prognosefähigkeit ein Kriterium für die Brauchbarkeit von Theorien. (Vgl Instrumentalismus von M. Friedman)

Wir unterscheiden verschiedene Arten von Prognosen, und versuchen die Prognosegüte zu messen.

Prognosebeispiel

Konsumfunktion für Ö, 1977-2005

Die OLS Schätzung für Österreich 1977 – 2005 ergibt, $n = 29$,

$$\hat{C}_t = -1771 + 0.408 Y_t^d + 0.576 C_{t-1} \quad R^2 = 0.996$$

(1457) (0.082) (0.086) $RSS = 32.6 \cdot 10^6$

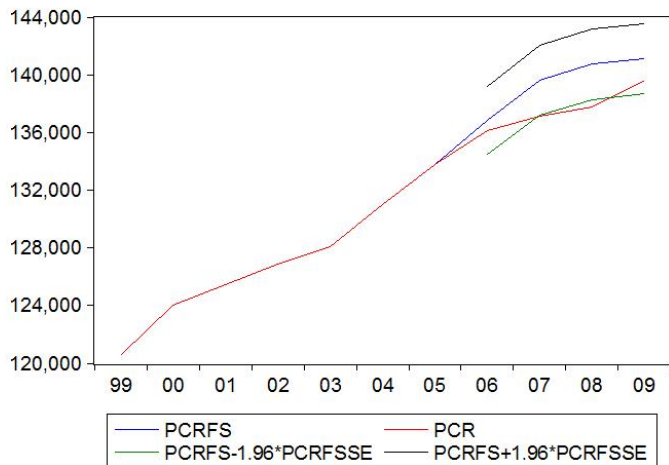
Wir haben unsere Konsumfunktion für die Periode 1977 – 2005 geschätzt und verwenden dieses Modell zur Prognose für die Periode von 2006 bis 2009.

Die Prognosefunktion der *ex post 1-Schritt-Prognose* ist für 2009, gegeben die Daten bis inklusive 2008,

$$\hat{C}_{2009} = -1771 + 0.408 Y_{2009}^d + 0.576 C_{2008}$$

Vgl die bedingte Erwartung aus Kap 4!

Prognose des privaten Konsums für Ö, 2006 – 2009



PCR ... tatsächlich beobachteter Verlauf

PCRFS ... Prognosewerte für PCR

PCRFS \pm 1.96 PCRFSSE ... 95% Prognoseintervall

Punktprognose

Punktprognose

Unsere theoretischen bzw geschätzten Modelle sind

$$y = \beta_0 + \beta_1 x + u \quad \text{bzw} \quad \hat{y} = b_0 + b_1 x$$

Unter einer **Punktprognose** verstehen wir den durchschnittlichen Wert von y für einen gegebenen Wert der Variablen x , x_0 . In unserem Modell ist das

$$(\hat{y}|x = x_0) = \hat{y}_0.$$

$$\hat{y}_0 = b_0 + b_1 x_0$$

Wir unterscheiden zwei Punktprognosen:

- ▶ in-sample Prognose und
- ▶ out-of-sample Prognose.

Punktprognose: in-sample

Für **in-sample** Prognosen werden x -Werte eingesetzt, die in der Stichprobe selbst auftreten.

► ZB für Querschnittsdaten:

Sie beobachten 10000 Haushalte, HH, aus dem Mikrozensus und messen deren monatlichen Konsumausgaben, C , und HH-Einkommen, Y . Sie wählen den HH Nr.157:

$$\hat{C}_{157} = b_0 + b_1 Y_{157}$$

\hat{C}_{157} ist die in-sample Prognose für die Konsumausgaben des HH 157. Sie entspricht dem durchschnittlichen Wert für Konsumausgaben eines HHs mit $Y = Y_{157}$.

Voraussetzung ist stets, dass das Modell korrekt ist.

Punktprognose: in-sample

- ▶ ZB für Zeitreihen:
Sie untersuchen die Abhängigkeit der AI-Rate vom Wi-Wachstum für die Periode 1980 – 2010. Die in-sample Prognosen entsprechen den Werten auf der Regressionsgeraden. Für 2001

$$\widehat{\Delta u}_{2001} = b_0 + b_1 \rho_{2001}$$

Punktprognose: out-of-sample

Für die **out-of-sample** Prognose werden für die erklärenden Variablen out-of-sample Werte eingesetzt.

- ▶ ZB für Querschnittsdaten:

Für die Mikrozensus Daten wollen sie zB für ein Einkommen, das nicht in der Erhebung beobachtet wird, die zugehörigen (durchschnittlichen) Konsumausgaben prognostizieren.

- ▶ ZB für Zeitreihen:

Sie wollen wissen, wie hoch die AI-Rate im Jahr 2011 ist. Für deren Prognose benötigen sie allerdings den Wert ρ im Jahr 2011, ρ_{2011} .

- ▶ Ist der Wert für 2011, ρ_{2011} , schon bekannt, handelt es sich um eine **ex post** Prognose.
- ▶ Ist der Wert für 2011, ρ_{2011} , nicht bekannt, so müssen sie ihn ebenfalls – zB mit Hilfe eines anderen Modells – prognostizieren. Dann handelt es sich um eine **ex ante** Prognose.

Punktprognose: 1-, 2-, mehr-Schritt-Prognosen, ZR

Zeitreihendaten liegen in einer zeitlichen Ordnung vor. Gegeben sind die Daten bis zum Zeitpunkt n . Die Prognose für den Zeitpunkt

- ▶ $(n + 1)$, heißt **1-Schritt-Prognose**,
- ▶ $(n + 2)$, heißt **2-Schritt-Prognose**,
- ▶ $(n + \tau)$, heißt **τ -Schritt-Prognose**.

Das maximal betrachtete τ heißt **Prognosehorizont**.

Bsp: In der obigen Konsumprognose ist der Horizont 4 Jahre (2006–2009). Es wurden jeweils *1-Schritt-ex post-Prognosen* berechnet: dh der Konsum des Folgejahres unter Verwendung des aktuellen Konsums und des *zukünftigen* Einkommens prognostiziert. Diese wurden für die Jahre 2006, 2007, 2008 und 2009 durchgeführt.

Prognosefehler

Prognosefehler

Der Prognosefehler ist

$$\hat{y}_0 - y_0$$

die Abweichung des prognostizierten Wertes, \hat{y}_0 , vom wahren, y_0 . Er kann positiv, oder negativ sein. Ist

- ▶ $\hat{y}_0 - y_0 > 0$ (positiv), so *überschätzen* wir, ist
- ▶ $\hat{y}_0 - y_0 < 0$ (negativ), so *unterschätzen* wir.

Analog dazu gibt es für Zeitreihendaten einen τ -Schritt-Prognosefehler.

$$\hat{y}_{n+\tau} - y_{n+\tau}$$

Erwartungswert des Prognosefehlers: $E(\hat{y}_0 - y_0) = 0$

Wir vergleichen die Prognoseregeln

$$\hat{y}_0 = b_0 + b_1 x_0$$

mit dem theoretischen Modellwert für $x = x_0$

$$y_0 = \beta_0 + \beta_1 x_0 + u_0$$

Beides setzen wir in die Definition des Prognosefehlers ein:

$$\hat{y}_0 - y_0 = (b_0 - \beta_0) + (b_1 - \beta_1)x_0 + u_0$$

Der **Erwartungswert des Prognosefehlers ist null**,

$$E(\hat{y}_0 - y_0) = 0$$

da die b_j erwartungstreu sind, $E(b_j - \beta_j) = 0$, und $E(u_0) = 0$.

Varianz des Prognosefehlers: $V(\hat{y}_0 - y_0)$

Die **Varianz des Prognosefehlers** ist (im bivariaten Modell)

$$V(\hat{y}_0 - y_0) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

Je weiter x_0 von \bar{x} entfernt ist (je untypischer x_0 ist), desto größer ist die Varianz.

$$S_{xx} = \sum_i (x_i - \bar{x})^2.$$

Wenn x_0 innerhalb des Wertebereichs der Beobachtungen liegt, können wird die Prognose auch in-sample Prognose nennen,
wenn x_0 außerhalb liegt, out-of-sample Prognose.

Exkurs: $V(\hat{y}_0 - y_0)$

$$\begin{aligned}V(\hat{y}_0 - y_0) &= V[(b_0 - \beta_0) + (b_1 - \beta_1)x_0 + u_0] = \\&= V(b_0 - \beta_0) + x_0^2 V(b_1 - \beta_1) + 2x_0 \text{Cov}(b_0 - \beta_0, b_1 - \beta_1) + V(u_0)\end{aligned}$$

Die Varianzen und Kovarianzen sind die Elemente von $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Für den bivariaten Fall eingesetzt

$$\begin{aligned}&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) + \sigma^2 \frac{x_0^2}{S_{xx}} - 2x_0 \sigma^2 \frac{\bar{x}}{S_{xx}} + \sigma^2 = \\&= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]\end{aligned}$$

Verteilung des Prognosefehlers

Unter der Annahme normal verteilter iid Fehler ist der Prognosefehler

$$(\hat{y}_0 - y_0) \text{ iid } N(0, \text{se}(\hat{y}_0)^2)$$

normal verteilt.

Der *Standardfehler des Prognosefehlers* ist $\text{se}(\hat{y}_0) = \sqrt{V(\hat{y}_0 - y_0)}$.

Prognoseintervall, Intervallprognose

Damit lassen sich zB 95% **Prognoseintervalle** konstruieren.

Sei unser Modell nun allgemein

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad \text{mit} \quad u \text{ iid } N(0, \sigma^2)$$

und die Werte $x_1^{(0)}, \dots, x_k^{(0)}$ gegeben, so lautet das Prognoseintervall mit einer Überdeckung von $(1 - \alpha)$

$$[\hat{y}_0 - t_{n-k-1, \alpha/2} \text{se}(\hat{y}_0), \hat{y}_0 + t_{n-k-1, 1-\alpha/2} \text{se}(\hat{y}_0)]$$

bzw für $(n - k - 1) \geq 30$ und $1 - \alpha = 0.95$

$$[\hat{y}_0 - 1.96 \text{se}(\hat{y}_0), \hat{y}_0 + 1.96 \text{se}(\hat{y}_0)]$$

Der *Unterschied zum Konfidenzintervall* besteht darin, dass ein KI einen fixen aber unbekanntem *Parameter* überdeckt, ein Prognoseintervall hingegen *Realisationen* einer ZVen.

Prognosegüte

Maße für die Prognosegüte: ME, MSE

Maße für die Prognosegüte dienen zum Vergleich verschiedener Prognosen für ein und derselben Variablen y . Wir verwenden die Zeitreihen-Notation. T bezeichne die maximale Anzahl der Prognoseschritte.

Mean error (mittlerer Fehler), **ME**:

$$ME = \frac{1}{T} \sum_{\tau=1}^T (\hat{y}_{n+\tau} - y_{n+\tau})$$

Der mittlere Fehler sollte null sein. Ist er nicht null, so sind die Prognosen verzerrt.

Mean square error (mittlerer quadratischer Fehler), **MSE**:

$$MSE = \frac{1}{T} \sum_{\tau=1}^T (\hat{y}_{n+\tau} - y_{n+\tau})^2$$

Er ist in der Regel größer als die Fehlervarianz. $MSE = bias^2 + Varianz$

Maße für die Prognosegüte: RMSE, MAD

Der **root mean square error, RMSE**, ist die Wurzel aus dem MSE.

$$RMSE = \sqrt{MSE}$$

Der RMSE hat den Vorteil gegenüber dem MSE, dass er auf der Skala von y misst.

Mean absolute error, MAD:

$$MAD = \frac{1}{T} \sum_{\tau=1}^T |\hat{y}_{n+\tau} - y_{n+\tau}|$$

Der MAD hat den Vorteil, dass er *robust* ist. Er reagiert auf einzelne große Fehler weniger als der MSE.

Maße für die Prognosegüte: MAPE

Will man nicht die Fehler im Niveau sondern relative Fehler bzw Fehler in Prozent angeben, so wählt man den **mean absolute percentage error, MAPE**

$$MAPE = 100 \cdot \frac{1}{T} \sum_{\tau=1}^T \left| \frac{\hat{y}_{n+\tau} - y_{n+\tau}}{y_{n+\tau}} \right|$$

Absolut kleine Fehler liefern auch einen großen Beitrag, wenn y klein ist.

Referenzen

Hackl 8

Wooldridge 18.5

- 1 Verwenden sie ihre selbst erstellten Daten (5 Beobachtungen) aus Kapitel 2. Legen sie ein Workfile für 10 Beobachtungen an. Geben sie ihre Daten in EViews ein, erzeugen sie ein Streudiagramm und berechnen sie die Regressionsgerade bzw verwenden sie `K5Ue1_R.txt`.
 - (a) Führen sie händisch in-sample Prognosen für $i = 3, 5$ durch, und berechnen sie den MSE.
 - (b) Wie groß ist der MSE, wenn sie alle in-sample Prognosen, $i = 1, \dots, 5$ bewerten.
 - (c) Wählen sie ein 6-tes Datenpaar. Berechnen sie händisch die out-of-sample Prognose und den Prognosefehler.

Übungen

- 1 (d) Vergleichen sie die Prognose mit dem wahren Wert. Überschätzen oder unterschätzen sie.
(e) Füllen sie die Positionen 6-10 der x Variablen mit
(e.1) ähnlichen Zahlen auf, die sie schon gewählt haben,
(e.2) Zahlen auf, die sich um Dimensionen unterscheiden,
und führen sie eine Prognose durch.
Vergleichen sie die Prognoseintervalle aus (e.1) und (e.2).

EViews:

Generate $\rightarrow x=0$ und $y=0$.

Markieren sie x und y . Rechtsklick `Open as Group`

\rightarrow `Edit +/-` Eingabe der Werte. Mit `Edit +/-` abschließen.

Analog sind Einträge und Änderungen der Daten durchzuführen.

- 2 Angenommen sie beobachten für Ungarn eine `income_inequality` von 4.1. Wie lautet ihre Prognose für die Anzahl der Gefängnisinsassen pro 100000 Einwohner.

[Anzahl d Häftlinge/100T Ew = $\exp(\text{imprisonment_log})$]

Verwenden sie die Ergebnisse aus dem Datenfile

`international_inequality.wf1:imprisonment_log` regressiert auf `income_inequality` bzw `international_inequality_K5Ue2_R.txt`.

Anleitung:

Entweder sie berechnen die Punktprognose mit der Hand nachdem sie die geschätzte Gleichung angeschrieben haben. Das Prognoseintervall ist allerdings schwieriger zu ermitteln.

Oder sie erweitern den Datenbereich des EViews-Files. Dazu rechts-klicken sie auf 23 im Range-Feld, und tragen für `observations` zB 24 ein.

Eingabe des neuen Wertes für Ungarn in Position 24 in `income_inequality` wie unter Bsp 1.

Das Modell wird nur für das `sample 1 23` geschätzt. Die Prognose (Forecast im Estimation output klicken) wird für den Bereich 24 24 durchgeführt.

- 3 Wählen sie das Datenfile `dat_s_01.wf1` bzw `dat_s_01_22_R.txt`. Berechnen sie für die Periode 1977 bis 2004 bzw 1977 bis 2010 die Regression `Quick` → `Estimate equation` `pcr c pyr pcr(-1)` bzw `cpr c yhr cpr(-1)`.
- (a) Führen sie eine in-sample Prognose durch. Wählen sie dazu im EViews-Forecast-Fenster `Static forecast`. Damit werden bei gegebenem Modell 1-Schritt-Prognosen berechnet.
- (b) Führen sie eine ex post Prognose für die Periode 2005 bis 2009 bzw 2011 bis 2020 durch. Wählen sie dazu ebenfalls `Static forecast`. Geben sie die Prognoseintervalle an.

- 3 (c) Erzeugen sie die Graphik zur Prognose des Konsums für Ö, 2005 – 2009 bzw 2011 – 2020, analog zu Folie 4.
- (d) Modifizieren sie die Konsumfunktion und schätzen sie $p_{cr} c_{pyr}$
 $p_{cr}(-1) p_{cr}(-2)$ bzw $c_{pr} c_{yhr} w_{lr}(-1)$.
- Wiederholen sie das Prognose-Experiment von (b) und vergleichen sie die RMSE und MAPE aus (b) und (c).
- (e) Überlegen sie sich wie man hier eine ex ante Prognose ermitteln kann.

EViews Befehle

Erweitern des Datenbereichs eines Workfiles:

Rechts-Klick auf den `Range` im Workfile unterhalb des Menüs. Ändern der gewünschten Werte.

Prognose:

Im `Estimation Fenster` → `Forecast` Prognoseperiode angeben.

Vergleich Prognose mit tatsächlichem Verlauf:

Siehe in `EviewsBefehle.pdf` unter `Prognose mit LS, ... und Plotten` einer Prognose.