Econometria
Tema 3: MRLM: Validació i Predicció

Ramon Alemany

Grau Estadística UB-UPC

Curs 2017-18

Presentació

Bibliografia

- GREENE, W. (1999)
  Análisis econométrico. 3a Ed.
  Capítol 6

- WOOLDRIDGE, J. (2009)
  Introducción a la Econometría. Un enfoque moderno. 4a Ed.
  Capítol 3

- STOCK, J. & WATSON, M. (2012)
  Introducción a la Econometría. 3a Ed.
  Capítol 6

Validació del model

1. Validació del Model

Els aspectes claus que caldrà analitzar al llarg d'aquesta tercera etapa són:

- La qualitat o bondat de l'ajust
- La significació dels paràmetres

Qualitat o bondat de l'ajust

Sempre que el model tingui terme independent, es pot afirmar que:

$$\mathrm{SQT} = \mathrm{SQR} + \mathrm{SQE}$$

Validació del model

Bondat de l'ajust

SQT = Suma dels Quadrats Total

$$\text{SQT} = \sum_{i=1}^{N} (y_i - \bar{y})^2 = Y'Y - N\bar{Y}^2$$

La SQT és la suma dels quadrats de les desviacions de la variable endògena respecte a la seva mitjana aritmètica.

SQR = Suma dels Quadrats de la Regressió

$$\text{SQR} = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2 = \hat{Y}'\hat{Y} - N\bar{Y}^2$$

La SQR és la suma dels quadrats de les desviacions dels valors estimats de l'endògena respecte a la seva mitjana.

Validació del model

Bondat de l'ajust

Aquest resultat s'obté suposant que la mitjana de la variable endògena és igual a la mitjana de la variable endògena estimada. Això es dóna sempre que el model tingui terme independent.

SQE = Suma dels Quadrats dels Errors

$$\text{SQE} = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} e_i^2 = e'e$$

Validació del model

Bondat de l'ajust

Donat que la SQT és el numerador de la variància mostral de la variable endògena, podem prendre la SQT com una mesura de la variabilitat total o dispersió de la variable endògena.

D'aquesta forma, la SQR es pot entendre com la part de la variabilitat total de la variable endògena que es veu explicada pel model estimat mentre que la SQE s'entén com la part de la variabilitat total de la variable endògena que no es veu explicada pel model estimat.

Validació del model

Bondat de l'ajust

Tenint això en compte, una primera mesura de bondat de l'ajust és el **Coeficient de Determinació**, R^2:

$$R^2 = 1 - \frac{\mathsf{SQE}}{\mathsf{SQT}}$$

En cas que el model tingui constant, el coeficient de determinació també es pot calcular:

$$R^2 = 1 - \frac{\mathsf{SQE}}{\mathsf{SQT}} = \frac{\mathsf{SQR}}{\mathsf{SQT}}$$

El coeficient R^2 mesura quina proporció de la variabilitat total de la variable endògena és explicada pel model. Així, l'ajust del model serà millor com més gran sigui el valor del coeficient de determinació.

Validació del model

Bondat de l'ajust

A més, el coeficient de determinació es troba acotat (sempre que el model tingui terme independent):

$$\begin{array}{c} \text{SQT} = \text{SQE} \\ \text{SQR} = 0 \end{array} \Longleftarrow \boxed{0 \leq R^2 \leq 1} \Longrightarrow \begin{array}{c} \text{SQT} = \text{SQR} \\ \text{SQE} = 0 \end{array}$$

Per tant, com més proper a 1 (0) sigui, millor (pitjor) serà l'ajust del model.

Però si el model no té terme independent:

$$-\infty \leq R^2 \leq 1$$

Validació del model

Bondat de l'ajust

També podem calcular el R^2 com el coeficient de correlació lineal al quadrat entre els valors observats de la variable dependent (y_i) i els valors predits (o ajustats) de la variable dependent $(\hat{y}_i)$:

$$R^2 = r^2_{y,\hat{y}} = \frac{\left[\sum (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right]^2}{\left[\sum (y_i - \bar{y})^2\right]\left[\sum (\hat{y}_i - \bar{\hat{y}})^2\right]}$$

Validació del model

Bondat de l'ajust

Malgrat la utilitat del coeficient de determinació, aquest presenta un problema quan el volem emprar per tal de seleccionar el millor model entre un conjunt de models ennierats.

Sempre acabarà seleccionant el model més general de tots atès que, a mesura que incorporem variables explicatives, la SQE disminueix.

Per tant, el coeficient de determinació no ha de ser utilitzat com a criteri per seleccionar entre models ennierats.

Validació del model

Bondat de l'ajust

Així, suposem que hem estimat els dos models següents:

$$\text{Model 1:} \quad y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

$$\text{Model 2:} \quad y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$$

els quals es troben ennierats entre ells.

En aquest cas, per poca capacitat explicativa que tingués la variable x_4, el coeficient de determinació del model 2 no seria mai inferior al del model 1.

Malgrat això, el model 2 presenta menys graus de llibertat que el model 1.

Validació del model

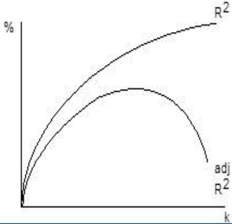
Bondat de l'ajust

Ens trobem doncs amb un problema de *trade off*:

Quantes més variables explicatives introduïm, més gran serà la capacitat explicativa del model però tindrem una pèrdua d'eficiència al haver de fer l'estimació amb menys graus de llibertat.

Això ho soluciona una mesura alternativa, el **Coeficient de Determinació Corregit**, $\bar{R}^2$:

$$\bar{R}^2 = 1 - \left[\frac{N-1}{N-k} (1 - R^2) \right]$$



Validació del model

Bondat de l'ajust

Quan incloguem una nova variable explicativa aleshores $(N-k)$ disminuirà mentre que el R^2 pot augmentar.

El seu rang de variació és:

$$0 \leq \bar{R}^2 \leq 1$$

i no es veurà afectat per l'augment en el nombre de variables explicatives.

Validació del model

Bondat de l'ajust

Consideracions en la interpretació del R^2:

- Un R^2 pot ser baix si el nombre d'observacions és reduït i alguns residus són grans.
- Un nombre reduït de graus de llibertat pot generar un R^2 elevat sense que la relació entre X i Y sigui estreta.
- El R^2 acostuma a ser reduït en models estimats amb mostres de tall transversal àmplies (mides mostrals superiors a 1000 observacions).
- També acostuma a ser reduït en models estimats en *variacions* (o en diferències) amb mostres de sèries temporals
- Amb dades temporals, si X i Y tenen una tendència comuna aleshores el R^2 pot ser artificialment elevat (regressió espúria).

Validació del model

Bondat de l'ajust

Consideracions en la interpretació del R^2:

- El R^2 depèn de la transformació de variables utilitzada. No és el mateix quan en el model expliquem Y que quan expliquem $\ln Y$, o en la regressió que explica Y que en la regressió que explica les variacions temporals de Y $(\triangle Y_t)$
- Que el R^2 sigui baix no indica que la variable dependent i les variables explicatives siguin estadísticament independents.
- Un R^2 reduït només indica feblesa en la relació lineal entre X i Y.
- Si X i Y tenen una relació estreta però no lineal la regressió pot tenir un R^2 reduït.

Validació del model

Significació dels paràmetres

Significació dels paràmetres

a) Significació econòmica
b) Significació estadística

Significació econòmica

Caldrà respondre a les preguntes següents:

1º) Quina **interpretació** tenen els valors de les estimacions dels paràmetres? Són coherents els signes dels paràmetres estimats (d'acord amb els preceptes de la teoria econòmica? Són coherents les seves magnituds?

Validació del model

Significació econòmica: interpretació

El que primer cal fer és veure com està expressat el model respecte els valors de les variables. Els dos casos més comuns són quan les variables estiguin mesurades en nivells o quan ho estiguin en logaritmes.

Si el model està especificat **en nivells**, aleshores els paràmetres reflecteixen l'efecte que tenen, en promig, les variacions unitàries de les variables explicatives sobre la variable endògena:

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

$$\beta_j = \frac{\partial y_i}{\partial x_j} \qquad j = 2, 3, \ldots, k$$

Validació del model

Significació econòmica: interpretació

Si el model està especificat **en logaritmes** neperians, la interpretació dels paràmetres és la d'una elasticitat:

$$Y_i = A \cdot X_{2i}^{\beta_2} \cdot X_{3i}^{\beta_3} \cdot \cdots \cdot X_{ki}^{\beta_k} \cdot u_i$$

Linealitzem prenent logaritmes:

$$\ln Y_i = \ln A + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ldots + \beta_k \ln X_{ki} + \ln u_i$$

$$\beta_j = \frac{\partial \ln Y}{\partial \ln X_j} = \frac{\partial Y}{\partial X_j} \cdot \frac{X_j}{Y} \qquad j = 2, 3, \ldots, k$$

Validació del model

Significació econòmica: interpretació

D'aquesta manera, la interpretació pot variar segons el tipus d'especificació del model.

A més, cal comprovar que les estimacions obtingudes siguin **coherents en signe i magnitud** amb la teoria econòmica. Si no és així, caldrà revisar la base de dades i el model utilitzat.

Validació del model

Significació econòmica: influència

2ª) Quina variable explicativa té més **influència** sobre la variable endògena?

De vegades pot ser interessant determinar quina variable explicativa té més influència sobre el comportament de la variable endògena. Però, les variables explicatives poden estar mesurades en diferents unitats i seria erroni identificar una variable com la més influent a partir del valor del paràmetre estimat.

Validació del model

Significació econòmica: influència

Si les variables explicatives estan mesurades en diferents unitats hi ha dues maneres d'esbrinar quina variable té major influència:

1. **Contribució mitjana d'una variable.**

   La contribució mitjana de la variable X_j és $\hat{\beta}_j \bar{X}_j$

2. **Coeficient beta estandarditzat.**

   Hi han dues maneres de calcular-lo. Una primera és a partir de l'expressió:

$$\hat{\beta}_j^* = \hat{\beta}_j \frac{S_{X_j}}{S_Y}$$

   on S_{X_j}, S_Y són les respectives desviacions típiques.

Validació del model

Significació econòmica: influència

Una altra manera és transformar el model dividint per la desviació típica de cada variable:

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

$$\frac{y_i}{S_y} = \beta_2^* \frac{x_{2i}}{S_{x_2}} + \ldots + \beta_k^* \frac{x_{ki}}{S_{x_k}} + u_i^*$$

I estimar aquest model transformat.

Els paràmetres estimats d'aquest model seran els coeficients beta estandarditzats del model original.

Validació del model

Significació estadística: significació individual

Significació estadística

Cal analitzar, mitjançant contrastos, la significació estadística individual i conjunta dels paràmetres del model.

1. Contrast de significació individual:

   $$H_0: \ \beta_j = 0$$
   $$H_a: \ \beta_j \neq 0$$

   L'estadístic de prova és:

$$\mathrm{t}_0 = \frac{\hat{\beta}_j}{\sqrt{\widehat{Var(\hat{\beta}_j)}}} \sim t_{N-k;\alpha/2}$$

Validació del model

Significació estadística: significació individual

El criteri de decisió que s'utilitza és el següent:

- Si $\mid \mathrm{t}_0 \mid \geq t_{N-k;\alpha/2}$ rebutgem la H_0.
  La conclusió és que, en base a l'evidència empírica, considerem el paràmetre β_j com estadísticament diferent de 0.
- Si $\mid \mathrm{t}_0 \mid < t_{N-k;\alpha/2}$ no rebutgem la H_0.
  La conclusió és que, en base a l'evidència empírica, considerem el paràmetre β_j com estadísticament igual a 0.

Validació del model

Significació estadística: significació conjunta

2. <u>Contrast de significació conjunta:</u>

$$\begin{array}{l} H_0: \; \beta_2 = \beta_3 = \ldots = \beta_k = 0 \\ H_a: \; \exists \beta_j \neq 0 \end{array}$$

L'estadístic de prova és:

$$\mathsf{F}_0 = \frac{\frac{\mathsf{SQR}}{k-1}}{\frac{\mathsf{SQE}}{N-k}} = \frac{\frac{\hat{\beta}' X' Y - N\bar{Y}^2}{k-1}}{\frac{Y'Y - \hat{\beta}' X' Y}{N-k}} \sim F_{k-1, N-k; \alpha}$$

Validació del model

Significació estadística: significació conjunta

El criteri de decisió que s'utilitza és el següent:

- Si $\mathsf{F}_0 \geq F_{k-1,N-k;\alpha}$ rebutgem la H_0.
  La conclusió és que, en base a l'evidència empírica, considerem el model com a globalment significatiu.
- Si $\mathsf{F}_0 < F_{k-1,N-k;\alpha}$ no rebutgem la H_0.
  La conclusió és que, en base a l'evidència empírica, considerem que el model no és globalment significatiu.

Contrast de Restriccions Lineals

2. Contrast de Restriccions Lineals

De vegades, a més de voler contrastar la significació estadística individual o conjunta dels paràmetres d'un MRLM, podríem estar interessats en contrastar determinats supòsits alternatius postulats per la Teoria Econòmica i que es manifesten en restriccions que han de complir els coeficients.

Per exemple: Funció de Producció Cobb-Douglas

$$\ln Q_i = \ln A + \beta_1 \ln K_i + \beta_2 \ln L_i + u_i$$

Contrast de Restriccions Lineals

Podríem estar interessats en contrastar si, per un conjunt d'economies analitzades, existeixen rendiments constants a escala, rendiments decreixents a escala o rendiments creixents a escala, és a dir:

$\beta_1 + \beta_2 = 1$ rendiments constants a escala
$\beta_1 + \beta_2 < 1$ rendiments decreixents a escala
$\beta_1 + \beta_2 > 1$ rendiments creixents a escala

Aquests contrastos on s'analitzen combinacions lineals de coeficients s'anomenen contrastos de restriccions lineals.

Contrast de Restriccions Lineals

Generalitzant:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + u_i$$

es poden contrastar tres tipus de restriccions:

a) **Restriccions lineals exactes d'igualtat:**
$a\beta_j + b\beta_l = c$
Si $c = 0 \longrightarrow$ Restricció lineal d'igualtat homogènia.
Si $c \neq 0 \longrightarrow$ Restricció lineal d'igualtat no homogènia.

b) **Restriccions lineals exactes de desigualtat:**
$a\beta_j + b\beta_l > c$; $a\beta_j + b\beta_l < c$

c) **Restriccions estocàstiques:**
$a\beta_j + b\beta_l > c + v$ $\quad v$: Terme d'error aleatori

Únicament ens centrarem en les restriccions lineals exactes d'igualtat.

Contrast de Restriccions Lineals

Per realitzar un contrast de restriccions líneals s'han de seguir dues etapes:

1) Formulació matricial de les restriccions lineals

En primer lloc, caldrà expressar les restriccions de forma matricial:

$$R_{qxk}\beta_{kx1} = r_{qx1}$$

on:

- q és el nombre de restriccions, $q \leq k$;
- R és una matriu constituïda pels coeficients fixos associats als paràmetres del vector de coeficients β;
- r és un vector que incorpora els termes independents de les restriccions lineals.

Contrast de Restriccions Lineals

Exemple:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i$$

on es vol contrastar si: $\quad 2\beta_2 - 4\beta_5 = 1$

Aleshores:

$$\underbrace{\begin{bmatrix} 0 & 2 & 0 & 0 & -4 \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}}_{\beta} = \underbrace{[1]}_{r}$$

Contrast de Restriccions Lineals

Exemple:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i$$

on es vol contrastar si:

$$\begin{aligned} \beta_2 &= 0 \\ \beta_3 - 2\beta_4 + 3\beta_5 &= -4 \\ \beta_5 &= -1 \end{aligned}$$

Aleshores:

$$\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{R} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}}_{\beta} = \underbrace{\begin{bmatrix} 0 \\ -4 \\ -1 \end{bmatrix}}_{r}$$

Contrast de Restriccions Lineals

2) Construcció de l'estadístic de prova

Es poden construir dos estadístics de prova alternatius:

a) Estadístic de prova basat en la matriu R i els vectors $\hat{\beta}_{MQO}$ i r.

b) Estadístic de prova basat en la comparació de la suma dels quadrats dels errors del model restringit i del model ampliat.

Contrast de Restriccions Lineals

a) **Estadístic de prova basat en la matriu R i els vectors $\hat{\beta}_{MQO}$ i r**

Les hipòtesis nul·la i alternativa són:

$H_0: \; R\beta = r$ (les restriccions són certes)

$H_a: \; R\beta \neq r$ (les restriccions no són certes)

Seguidament, es pot construir l'estadístic de prova com:

$$\mathrm{F}_0 = \frac{(R\hat{\beta} - r)' \left[R(X'X)^{-1}R'\right]^{-1} (R\hat{\beta} - r)/q}{e'e/N-k} \sim F_{q,N-k;\alpha}$$

Contrast de Restriccions Lineals

b) **Estadístic de prova basat en la en la comparació de la suma dels quadrats dels errors del model restringit i del model ampliat**

Les hipòtesis nul·la i alternativa són:

$H_0: \ R\beta = r$ (les restriccions són certes)

$H_a: \ R\beta \neq r$ (les restriccions no són certes)

Seguidament, es pot construir l'estadístic de prova com:

$$\mathsf{F}_0 = \frac{(\mathsf{SQE}_R - \mathsf{SQE}_A)/q}{\mathsf{SQE}_A/N-k} \sim F_{q,N-k;\alpha}$$

Contrast de Restriccions Lineals

on:

SQE_R és la suma dels quadrats dels errors d'estimació MQO del model restringit (és a dir, del model que incorpora les restriccions) i

SQE_A és la suma dels quadrats dels errors d'estimació MQO del model ampliat (és a dir, del model que no incorpora les restriccions).

Contrast de Restriccions Lineals

El criteri de decisió que s'utilitza en ambdós casos és el següent:

- Si $\mathrm{F}_0 \geq F_{q,N-k;\alpha}$ rebutgem la H_0.

  Es rebutja la hipòtesi nul·la i, per tant, les restriccions no són certes en l'àmbit de la població.

- Si $\mathrm{F}_0 < F_{q,N-k;\alpha}$ no rebutgem la H_0.

  No es rebutja la hipòtesi nul·la i, per tant, les restriccions són certes en l'àmbit de la població.

Contrast de Restriccions Lineals

Si el model té terme independent aleshores l'estadístic de prova del contrast de restriccions es pot reformular a partir dels coeficients de determinació del model ampliat i del model restringit:

$$\mathsf{F}_0 = \frac{(R_A^2 - R_R^2)/q}{(1 - R_A^2)/N - k} \sim F_{q,N-k;\alpha}$$

El contrast de significació individual d'un paràmetre i el contrast de significació global del model són casos particulars del contrast de restriccions lineals

Predicció puntual i per interval

3. Predicció puntual i per interval. Valoració de les prediccions

Un model ja estimat i validat es pot utilitzar, entre d'altres objectius, per fer prediccions.

Si tenim un model amb dades de sèrie temporal podem estar interessats en predir el comportament de la variable endògena pel futur.

Si tenim un model amb dades de tall transversal podem estar interessats en predir el comportament de la variable endògena per a un individu que no està a la mostra.

Predicció puntual i per interval

Es poden fer dos tipus de prediccions:

- Prediccions Puntuals: es tracta de predir un valor.
- Prediccions per Interval: es tracta de determinar un interval de possibles valors entre els quals estigui el valor de la variable endògena.

Per fer les prediccions haurem de suposar que les hipòtesis bàsiques es mantenen per les observacions extra-mostrals.

És especialment important que es compleixi la hipòtesi de permanència estructural. En cas contrari, no tindria sentit fer una predicció.

Predicció puntual i per interval

Predicció Puntual:

Suposem que la variable endògena ajustada per a una observació i és:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \ldots + \hat{\beta}_k x_{ki}$$

Si volem predir el valor de la variable endògena per a una observació $N+h$ utilitzarem:

$$\hat{y}_{N+h} = \hat{\beta}_1 + \hat{\beta}_2 x_{2N+h} + \hat{\beta}_3 x_{3N+h} + \ldots + \hat{\beta}_k x_{kN+h}$$

Com es pot veure, per fer una predicció puntual cal conèixer els valors de les variables explicatives per l'observació $N+h$.

Predicció puntual i per interval

Predicció per interval:

Per obtenir la predicció per interval sobre la variable endògena per a l'observació $N+h$, s'utilitza la següent expressió:

$$P\left(|y_{N+h}-\hat{y}_{N+h}| \leq t_{N-k;\alpha/2}\sqrt{\hat{\sigma}_u^2\left(1+x'_{N+h}(x'x)^{-1}x_{N+h}\right)}\right)=1-\alpha$$

Un cop que es tria el nivell de significació α, la probabilitat que el valor de la variable endògena per a l'observació $N+h$ es trobi en el interval:

$$\hat{y}_{N+h} \pm t_{N-k;\alpha/2}\sqrt{\hat{\sigma}_u^2\left(1+x'_{N+h}(x'x)^{-1}x_{N+h}\right)}$$

és $1-\alpha$.

Predicció puntual i per interval

Valoració de les prediccions

Valoració de les prediccions:

Per tal de valorar les prediccions que es fan en un model de regressió lineal múltiple existeixen diversos instruments.

Si definim l'error de predicció com:

$$e_N(h) = y_{N+h} - \hat{y}_N(h)$$

podem formular alguns d'aquests instruments de valoració de les prediccions:

Predicció puntual i per interval

Valoració de les prediccions

- **Error Absolut Mig (EAM):**

$$\text{EAM}(H) = \frac{1}{H}\sum_{h=1}^{H} \mid e_N(h) \mid = \frac{1}{H}\sum_{h=1}^{H} \mid y_{N+h} - \hat{y}_N(h) \mid$$

- **Error Quadràtic Mig (EQM):**

$$\text{EQM}(H) = \frac{1}{H}\sum_{h=1}^{H} e_N^2(h) = \frac{1}{H}\sum_{h=1}^{H} (y_{N+h} - \hat{y}_N(h))^2$$

De vegades es calcula l'arrel quadrada de l'EQM.

Predicció puntual i per interval

Valoració de les prediccions

Per tal que un conjunt de prediccions siguin bones cal que l'EAM i l'EQM siguin el més petits possibles. Però aquests instruments tenen la limitació de dependre de les unitats de mesura de la variable endògena. És a dir, són mesures no adimensionals.

Per tant, un valor concret d'aquests instruments no ens informa realment sobre si les prediccions són bones o no ho són. La utilitat d'aquestes eines es limita a la seva comparació.

Davant de models alternatius, aquell que presenti un EAM o EQM de predicció inferior serà el que farà millors prediccions. Però en cap cas sabrem, a partir d'aquestes mesures, si les prediccions són bones.

Predicció puntual i per interval

Valoració de les prediccions

Per mesurar la capacitat predictiva d'un mètode concret necessitem d'un indicador que sigui adimensional:

- **Error Percentual Absolut Mig (EPAM):**

$$\text{EPAM}(H) = \frac{100}{H} \sum_{h=1}^{H} \frac{\mid e_N(h) \mid}{Y_{N+h}}$$

$\text{EPAM} \leq 1\,\%$: Molt bona capacitat predictiva
$1\,\% < \text{EPAM} \leq 3\,\%$: Bona capacitat predictiva
$3\,\% < \text{EPAM} \leq 5\,\%$: Capacitat predictiva regular
$5\,\% < \text{EPAM}$: Baixa/molt baixa capacitat predictiva

Predicció puntual i per interval

Valoració de les prediccions

Una altra mesura per tal d'avaluar la capacitat predictiva d'un model és el:

- **Coeficient de desigualtat de Theil:**

$$\text{CDT} = \frac{\sqrt{\frac{\sum(\hat{Y}_i - Y_i)^2}{H}}}{\sqrt{\frac{\sum \hat{Y}_i^2}{H}} + \sqrt{\frac{\sum Y_i^2}{H}}}$$

El CDT pot presentar valors compresos entre 0 i 1, sent el millor valor possible el més proper a 0.

Econometria
Tema 3: MRLM: Validació i Predicció

Ramon Alemany

Grau Estadística UB-UPC

Curs 2017-18