Formation « modèles de prévision »

# Comparaison des modèles de prévision


Intervenants

Morgane Glotain
Alain Quartier-la-Tente


29-31 mars 2017

# Introduction

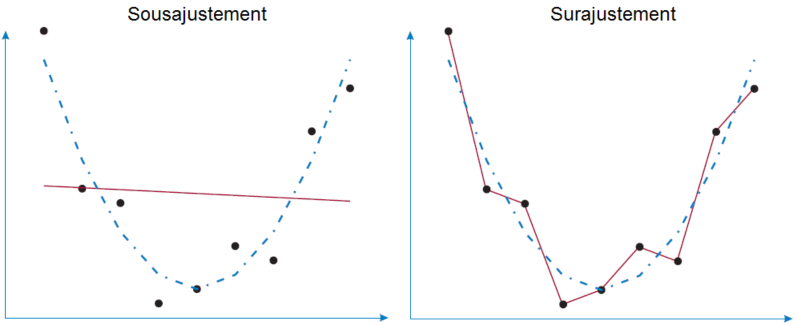
Lorsque l'on fait des estimations de modèles il faut deux échantillons :

- **évaluation** pour «  trouver » le modèle $\rightarrow$ capacité du modèle à expliquer le passé en ayant connaissance de toutes les données
- **test** pour tester la «  robustesse » du modèle $\rightarrow$ capacité du modèle s'adapter aux nouvelles informations et qualité lors des exercices de prévision

Problème cas temporel : pas beaucoup de recul temporel, comment créer les deux échantillons ?

# Introduction

Éviter le problème de surajustement (échantillon d'évaluation vs échantillon test)




→ Il faut faire un compromis entre biais (sousajustement) et variance (surajustement)

# Sommaire

# Les prévisions dans l'échantillon

Les prévisions **dans l'échantillon** (*in sample*) sont les prévisions « classiques » : on estime les coefficients sur toute la période connue et on calcule la prévision sur toute cette période

- ✓ la recherche de modèles est effectuée avec les prévisions dans l'échantillon
- ✓ permet de mesurer la dispersion des erreurs de prévision
- ✗ ne permet pas de voir si le modèle aurait bien marché lors des périodes de crises (quelle aurait été la prévision du modèle ?)
- ✗ peut masquer effets de surajustement (ajustement artificiellement bon mais capacités prédictives faibles)

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
$\rightarrow$ on parle aussi de *leave-one-out cross-validation* (LOOCV)

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
$\rightarrow$ on parle aussi de *leave-one-out cross-validation* (LOOCV)

Exemple si le modèle commence en 2000T1 :

| |
|---|
| 2000T1 |
| $\vdots$ |
| 2009T4 |

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
$\rightarrow$ on parle aussi de *leave-one-out cross-validation* (LOOCV)

Exemple si le modèle commence en 2000T1 :

| |
|---|
| 2000T1 |
| ⋮ |
| 2009T4 |

$\longrightarrow$ On estime les coefficients sur la période 2000T1-2009T4 pour prévoir le point de 2010T1

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
$\rightarrow$ on parle aussi de *leave-one-out cross-validation* (LOOCV)

Exemple si le modèle commence en 2000T1 :

| |
|---|
| 2000T1 |
| ⋮ |
| 2009T4 |
| 2010T1 |

$\longrightarrow$ On estime les coefficients sur la période 2000T1-2010T1 pour prévoir le point de 2010T2

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
→ on parle aussi de *leave-one-out cross-validation* (LOOCV)

Exemple si le modèle commence en 2000T1 :

| |
|---|
| 2000T1 |
| ⋮ |
| 2009T4 |
| 2010T1 |
| 2010T2 |

→ On estime les coefficients sur la période 2000T1-2010T2 pour prévoir le point de 2010T3

# Les prévisions en temps réel (1/2)

Les prévisions **en temps réel** sont les prévisions dynamiques obtenues en réactualisant les coefficients à chaque trimestre pour prévoir le trimestre suivant
→ on parle aussi de *leave-one-out cross-validation* (LOOCV)

Exemple si le modèle commence en 2000T1 :

| |
|---|
| 2000T1 |
| ⋮ |
| 2009T4 |
| 2010T1 |
| 2010T2 |
| 2010T3 |

→ On estime les coefficients sur la période 2000T1-2010T3 pour prévoir le point de 2010T4

# Les prévisions en temps réel (2/2)

- ✓ permet de mesurer la qualité prédictive du modèle
- ✓ permet d'étudier le comportement du modèle lors de certains exercices particuliers (parfois masqués dans la prévision dans l'échantillon par des indicatrices)
- ✗ le modèle qui aurait été retenu en 2010 n'est pas forcément le même (le jeu de régresseur n'est pas forcément optimal pour chaque exercice de prévision)
- ✗ en réalité c'est une prévision en pseudo-temps réel (indicateurs révisés)
- ✗ nécessite un recul temporel important pour la première estimation (10 ans ?)

La racine carrée de l'erreur quadratique moyenne – *root mean square error* – (RMSE) des prévision en temps réel est également appelée **statistique de validation croisée (LOOCV)**

# Sommaire

# Définition du RMSE

Comment comparer les prévisions dans l'échantillon ou en temps réel ? Il faut un indicateur quantitatif !

Indicateur retenu : le RMSE qu'il faut **minimiser**

$$RMSE = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(y_{t,obs} - y_{t,prev})^2}$$

# Définition du RMSE

Comment comparer les prévisions dans l'échantillon ou en temps réel ? Il faut un indicateur quantitatif !

Indicateur retenu : le RMSE qu'il faut **minimiser**

$$RMSE = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(y_{t,obs} - y_{t,prev})^2}$$

Si $RMSE > \sigma_y$ alors une prévision « naïve » (par la moyenne) est de meilleure qualité

D'autres indicateurs existent (ex : MAE) mais le RMSE donne un poids important aux valeurs extrêmes (et donc aux grandes erreurs de prévision)

# Le test de Diebold-Mariano

Généralement, plus $\sigma_y$ est grand plus le RMSE sera élevé. Il faut donc un test pour les comparer : c'est le test de **Diebold-Mariano** !

# Le test de Diebold-Mariano

Généralement, plus $\sigma_y$ est grand plus le RMSE sera élevé. Il faut donc un test pour les comparer : c'est le test de **Diebold-Mariano** !

Soient deux séries d'erreurs $(e_{1,t})_{t=1..T}$ et $(e_{2,t})_{t=1..T}$. Le test de Diebold-Mariano teste :

$$\begin{cases} (H_0) \; : \; \mathbb{E}[d_t] = 0 \\ (H_1) \; : \; \mathbb{E}[d_t] \neq 0 \text{ ou } \mathbb{E}[d_t] < 0 \text{ ou } \mathbb{E}[d_t] > 0 \end{cases} \quad \text{avec } d_t = g(e_{1,t}) - g(e_{2,t})$$

Avec $g$ une **fonction de perte** (par exemple $g\colon\ x \mapsto x^2$ pour RMSE)

# Le test de Diebold-Mariano

Généralement, plus $\sigma_y$ est grand plus le RMSE sera élevé. Il faut donc un test pour les comparer : c'est le test de **Diebold-Mariano** !

Soient deux séries d'erreurs $(e_{1,t})_{t=1..T}$ et $(e_{2,t})_{t=1..T}$. Le test de Diebold-Mariano teste :

$$\begin{cases} (H_0)\ :\ \mathbb{E}[d_t] = 0 \\ (H_1)\ :\ \mathbb{E}[d_t] \neq 0 \text{ ou } \mathbb{E}[d_t] < 0 \text{ ou } \mathbb{E}[d_t] > 0 \end{cases} \quad \text{avec } d_t = g(e_{1,t}) - g(e_{2,t})$$

Avec $g$ une **fonction de perte** (par exemple $g: x \mapsto x^2$ pour RMSE)

✓ test robuste aux hypothèses sur les résidus (valide si résidus autocorrélés, hétéroscédastiques et ne suivent pas une loi normale)

✓ indépendant de la fonction de perte (MAE, puissance 4, etc.)

⚠ On compare les prévisions mais pas les modèles (pas de meilleur modèle)

# Sommaire

# $R^2$ et $R_a^2$ (1/2)

Dans le modèle $y_i = \beta_0 + \beta X_i + \varepsilon_i$ on a :

$$R^2 = 1 - \frac{SCR}{SCT} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\text{var. expliquée par les résidus}}{\text{var. totale}}$$

Mais le $R^2 \nearrow$ en rajoutant une variable, même non significative. Le $R_a^2$ corrige du nombre de variables explicatives ($k$) :

$$R_a^2 = 1 - \frac{\text{var. expliquée par les résidus}}{\text{var. totale}} \frac{n-1}{n-k}$$

Le $R_a^2 \nearrow$ si une variable explicative significative est rajoutée

⚠ Chercher un modèle qui maximise le $R_a^2$ ne va pas permettre de trouver celui qui donne les meilleures prévisions et peut aboutir à des erreurs

# $R^2$ et $R_a^2$ (2/2)

1. Maximiser le $R_a^2$ n'est pas équivalent à minimiser les résidus (ex : étudier une série en différence peut $\searrow$ le $R_a^2$ mais conduire à une meilleure prévision)
2. Lorsque la variable à prévoir comporte une tendance (non stationnaire) : rajouter des retards va $\nearrow$ le $R_a^2$ mais $\nearrow$ l'erreur de prévision. En général : + la tendance est forte + le $R_a^2$ est fort mais expliquer la tendance n'améliore pas les prévisions
3. Utiliser des données non CVS va $\nearrow$ le $R_a^2$ mais uniquement parce que l'on « explique » les mouvements saisonniers
4. Expliquer une valeur extrême par une indicatrice va toujours $\nearrow$ le $R_a^2$ mais pas nécessairement améliorer les prévisions
5. « Torturer les données » pour améliorer légèrement le $R_a^2$ peut améliorer la prévision mais sûrement du fait d'évènements aléatoires qui ne se reproduiront pas lors des exercices de prévision

# Deux critères d'information : l'AIC et le BIC (1/3)

Critères d'information : mesure la qualité d'un modèle statistique en pénalisant le maximum de la vraisemblance par le nombre de variables explicatives

$$AIC = 2p - 2\log\mathcal{L} \qquad\qquad BIC = SC = p\log(N) - 2\log\mathcal{L}$$

Avec :

$$\begin{cases} p & = & \text{nombre de variables explicatives} \\ N & = & \text{nombre d'observations} \\ \mathcal{L} & = & \text{maximum de la fonction de vraisemblance} \end{cases}$$

Il faut les minimiser et comparer les modèles avec ces critères uniquement lorsqu'ils sont évalués sur le même échantillon

# Deux critères d'information : l'AIC et le BIC (2/3)

L'**AIC** :

✓ asymptotiquement équivalent à la statistique de validation croisée LOOCV (minimiser AIC $\Longleftrightarrow$ minimiser RMSE temps réel)

⚠ l'AIC n'est pas une mesure de la qualité des prévisions

✓ efficace (meilleur compromis biais-variance)

✗ a tendance à surajuster

✗ ne converge pas vers le « vrai modèle » (mais existe-t-il vraiment ?)

# Deux critères d'information : l'AIC et le BIC (2/3)

L'**AIC** :

✓ asymptotiquement équivalent à la statistique de validation croisée LOOCV (minimiser AIC $\iff$ minimiser RMSE temps réel)

⚠ l'AIC n'est pas une mesure de la qualité des prévisions

✓ efficace (meilleur compromis biais-variance)

✗ a tendance à surajuster

✗ ne converge pas vers le « vrai modèle » (mais existe-t-il vraiment ?)

Le **BIC** :

✓ converge vers le « vrai modèle » (consistant)

✗ a tendance à sousajuster

# Deux critères d'information : l'AIC et le BIC (2/3)

L'**AIC** :

- ✓ asymptotiquement équivalent à la statistique de validation croisée LOOCV (minimiser AIC $\Longleftrightarrow$ minimiser RMSE temps réel)

  l'AIC n'est pas une mesure de la qualité des prévisions
- ✓ efficace (meilleur compromis biais-variance)
- ✗ a tendance à surajuster
- ✗ ne converge pas vers le « vrai modèle » (mais existe-t-il vraiment ?)

$\Longrightarrow$ mieux pour un modèle prédictif

Le **BIC** :

- ✓ converge vers le « vrai modèle » (consistant)
- ✗ a tendance à sousajuster

$\Longrightarrow$ mieux pour un modèle prédictif

# Deux critères d'information : l'AIC et le BIC (3/3)

D'autres critères d'information existent, par exemple dans `gets` :

$$HQ = 2p\log(\log(N)) - 2\log\mathcal{L}$$

⚠ Les définitions peuvent changer entre les logiciels (constante multiplicative en plus, additive, etc.). Par exemple sous Eviews :

$$AIC = 2\frac{p}{N} - 2\frac{\log\mathcal{L}}{N}$$
$$BIC = \frac{p\log(N)}{N} - 2\frac{\log\mathcal{L}}{N}$$
$$HQ = 2\frac{p\log(\log(N))}{N} - 2\frac{\log\mathcal{L}}{N}$$

# Conclusion

Pour comparer la qualité des prévisions des étalonnages :

- utiliser les RMSE des prévisions dans l'échantillon et en temps réel (qu'il faut minimiser)
  $\rightarrow$ les comparer en utilisant le test de Dieblod-Mariano (sous R fonction `forecast::dm.test` en changeant éventuellement l'hypothèse alternative avec l'option `alternative`)
- pour la sélection de modèles préférer les critères d'information (notamment l'AIC) aux $R^2$ et $R_a^2$
- ne pas négliger l'analyse graphique : à quelle période les erreurs de prévisions sont importantes ? les retournements conjoncturels sont-ils captés ? les résidus sont-ils toujours de même signe ?

# Merci de votre attention !


Morgane Glotain morgane.glotain@insee.fr
Alain Quartier-la-Tente alain.quartier-la-tente@insee.fr