Les Prévisions tirées des processus ARMA(p,q)

Gilbert Colletaz

4 décembre 2018

Résumé

Dans cette partie nous allons discuter des caractéristiques générales afférentes aux prévisions de variables stationnaires. Par la suite, nous présenterons la méthode de construction des prévisions obtenues à partir de processus ARMA(p,q) ou SARMA(p,q)$\times$ARMA(P,Q)$_s$, ainsi que celle de leur écart-types. Enfin, nous donnerons les commandes exécutant les calculs en question sous SAS et R.

Table des matières

1 Les prévisions des processus stationnaires

1.1 espérance et variance conditionnelles tendent vers les moments non conditionnels

Soit X$_t$ une variable stationnaire ayant une écriture de Wold, $X_t = \mu_X + \sum_{i=0}^{\infty} \psi_i u_{t-i}$, où u est un processus en bruit blanc de variance σ_u^2 et $\phi_0 = 1$. Comme souvent pour simplifier les expressions, on va centrer cette variable, soit $x_t = X_t - \mu_X$. La variance non conditionnelle de cette variable est donnée par :

$$Var(x_t) = \sigma_u^2 \sum_{i=0}^{\infty} \psi_i^2 \tag{1}$$

En t, la prévision optimale de x à l'horizon h est :

$$\begin{aligned}
{}_t x_{t+h} &= E_t[x_{t+h}] \\
&= E_t\left[\sum_{i=0}^{\infty} \psi_i u_{t+h-i}\right] \\
&= \psi_h u_t + \psi_{h+1} u_{t-1} + \ldots
\end{aligned}$$

et, comme on l'a déjà vu, l'erreur de prévision qui lui est associée, $x_{t+h} -_t x_{t+h} = \sum_{i=0}^{h+1} \psi_i u_{t+h-i}$ est d'espérance conditionnelle nulle en t.

Ainsi,

$$Var_t(x_{t+h} -_t x_{t+h}) = \sigma_u^2 \sum_{i=0}^{h+1} \psi_i^2 \tag{2}$$

Si on fait tendre l'horizon de prévision vers l'infini, alors :

$$\begin{aligned}\lim_{h\to\infty} Var_t(x_{t+h} -_t x_{t+h}) &= \sigma_u^2 \sum_{i=0}^{\infty} \psi_i^2, \quad \text{et d'après (1)} \\ &= Var(x_t) \end{aligned} \tag{3}$$

Ce qui signifie que lorsque l'horizon de prévision augmente alors l'information disponible en t, date où on construit la prévision, finit par devenir inutile et la précision obtenue en se basant sur cette information, *i.e.* la variance conditionnelle, devient égale à la précision obtenue en se passant de cette information, *i.e.* la variance non conditionnelle.
On peut montrer que pour des séries stationnaires ergodiques, le même raisonnement vaut pour l'espérance. En effet, si on repart du dernier résultat, et en se rappelant que ${}_tx_{t+h}$ est un prédicteur sans biais de x_{+h}, on peut écrire :

$$\lim_{h\to\infty} Var_t(x_{t+h} - {}_tx_{t+h}) = Var(x_{t+h}) \tag{4}$$

$$\lim_{h\to\infty} E_t[(x_{t+h} - {}_tx_{t+h})^2] = E[(x_{t+h} - E[x_t])^2] \tag{5}$$

$$\Rightarrow \lim_{h\to\infty} {}_tx_{t+h} = E(x_t) \tag{6}$$

Lorsque $h \to \infty$ alors $E_t[x_{t+h}] = E[x_{t+h}] = 0$, soit encore, sur la variable non centrée, $\lim_{h\to\infty} E_t[X_{t+h}] = E[X_{t+h}] = \mu_X$: quand l'horizon de prévision augmente, la valeur de la prévision retourne vers l'espérance de la variable que l'on cherche à prévoir. Ce retour pour des processus stationnaires est de nature exponentielle : le temps mis par les prévisions pour rejoindre l'espérance non conditionnelle dépend de la taille des autocorrélations, mais sauf à avoir des corrélations proches de l'unité, il est généralement assez court. Pour cette raison, vous devez comprendre que les prévisions issues de modèles ARMA et SARMA sont surtout utiles pour construire des prévisions de court terme.

1.2 la précision se détériore avec l'horizon de prévision

Ce résultat, logique, résulte directement de l'équation (2). Si $h1$ et $h2$ sont deux horizons de prévisions tels que $h1 < h2$ alors

$$Var({}_tx_{t+h1}) < Var({}_tx_{t+h2}) < Var(x_t) \tag{7}$$

La largeur des intervalles de confiance construits autour des prévisions ponctuelles avec un seuil de risque donné augmente avec l'horizon et est bornée par une largeur asymptotique définie à partir de la variance non conditionnelle de la variable étudiée.

2 Les prévisions des processus ARMA

2.1 les prévisions itérées

En pratique, les prévisions construites à partir de processus ARMA estimés sont qualifiées de prévisions itérées simplement parce qu'on utilise, si besoin, dans l'équation de construction de la prévision à un horizon h les prévisions calculées antérieurement pour les horizons $h-1$, $h-2$,...,1.

On peut illustrer cette méthode avec quelques exemples simples mais qui s'étendent sans difficulté à des processus plus complexes :

- Cas d'un AR(1). Soit à l'issue de l'étape d'estimation le processus : $x_{t+1} = \hat{\phi}_0 + \hat{\phi}_1 x_t + \hat{u}_{t+1}$. L'information pertinente et connue en t pour prévoir x_{t+1} est constituée de $\{\hat{\phi}_0, \hat{\phi}_1, x_t\}$. En passant à l'espérance conditionnelle, On a :
  — pour l'horizon 1 :${}_t\hat{x}_{t+1} = \hat{\phi}_0 + \hat{\phi}_1 x_t$
  — pour l'horizon 2 : ${}_t\hat{x}_{t+2} = \hat{\phi}_0 + \hat{\phi}_1 E_t[x_{t+1}] = \hat{\phi}_0 + \hat{\phi}_1\, {}_t\hat{x}_{t+1}$
  — pour les horizons $h \geq 3$: ${}_t\hat{x}_{t+h} = \hat{\phi}_0 + \hat{\phi}_1\, {}_t\hat{x}_{t+h-1}$

- Cas d'un MA(2). Soit : $x_t = \hat{\mu} + \hat{u}_t - \hat{\theta}_1 \hat{u}_{t-1} - \hat{\theta}_2 \hat{u}_{t-2}$. L'information pertinente et connue en t pour prévoir x_{t+1} est constituée de $\{\hat{\mu}, \hat{\theta}_1, \hat{\theta}_2, \hat{u}_t \hat{u}_{t-1}, \hat{u}_{t-2}\}$. En passant à l'espérance conditionnelle, et sachant que les résidus empiriques sont supposés être orthogonaux entre eux, on obtient :
  — pour l'horizon 1 :${}_t\hat{x}_{t+1} = \hat{\mu} - \hat{\theta}_1 \hat{u}_t - \hat{\theta}_2 \hat{u}_{t-1}$
  — pour l'horizon 2 : ${}_t\hat{x}_{t+2} = \hat{\mu} - \hat{\theta}_2 \hat{u}_t$
  — pour les horizons $h \geq 3$: ${}_t\hat{x}_{t+h} = \hat{\mu}$

- Cas d'un ARMA(1,2). On part de l'équation empirique $x_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \hat{u}_t - \hat{\theta}_1 \hat{u}_{t-1} - \hat{\theta}_2 \hat{u}_{t-2}$. En reprenant la démarche des deux exemples précédents, il vient :
  — pour l'horizon 1 :${}_t\hat{x}_{t+1} = \hat{\phi}_0 + \hat{\phi}_1 x_t - \hat{\theta}_1 \hat{u}_t - \hat{\theta}_2 \hat{u}_{t-1}$
  — pour l'horizon 2 : ${}_t\hat{x}_{t+2} = \hat{\phi}_0 + \hat{\phi}_1\, {}_t\hat{x}_{t+1} - \hat{\theta}_2 \hat{u}_t$
  — pour les horizons $h \geq 3$: ${}_t\hat{x}_{t+h} = \hat{\phi}_0 + \hat{\phi}_1\, {}_t\hat{x}_{t+h-1}$

2.2 les intervalles de confiance associés aux prévisions itérées

Même si les ordres p et q retenus sont les bons, l'emploi d'un ARMA estimé pour construire des prévisions implique que celles-ci seront entachées de deux sources d'erreur : les innovations futures d'une part et la différences entre les vrais coefficients et leurs estimations d'autre part. Pour le montrer, on considère un processus AR(1) sur une variable x : $x_t = \phi_1 x_{t-1} + u_t$ et son estimation : $x_t = \hat{\phi}_1 x_{t-1} + \hat{u}_t$. On peut vérifier immédiatement que l'erreur de prévision à l'horizon d'une période est : $x_{t+1} -_t x_{t+1} = (\phi_1 - \hat{\phi}_1) x_t + u_{t+1}$.
En conséquence, la variance de ces erreurs devrait dépendre de la variance des coefficients estimés et de la variance des innovations. En pratique cependant tous les logiciels font l'hypothèse que le modèle utilisé est le vrai modèle. Avec cette conjecture qui, comme vous le savez, correspond à un événement de probabilité nulle, seule la variance des innovations va intervenir dans l'expression de la variance des prévisions et par ailleurs $\hat{u}_t = u_t$. Si on reprend, sous cette hypothèse, les exemples précédents, on vous laisse vérifier que les expressions des erreurs de prévision afférentes à des prédicteurs construits à une date t pour un horizon h, erreurs notées ${}_te_{t+h}$, sont les suivantes :

- Cas d'un AR(1). Soit donc : $x_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + u_t$, alors :
  — pour l'horizon 1 : ${}_te_{t+1} = x_{t+1} -_t \hat{x}_{t+1} = u_{t+1}$
  — pour l'horizon 2 : ${}_te_{t+2} = x_{t+2} -_t \hat{x}_{t+2} = u_{t+2} + \phi_1 u_{t+1}$
  — pour les horizons $h \geq 3$: ${}_te_{t+3} = x_{t+h} -_t \hat{x}_{t+h} = \sum_{i=1}^{h} \hat{\phi}_1^{h-i} u_{t+i}$

- Cas d'un MA(2). Avec $x_t = \hat{\mu} + u_t - \hat{\theta}_1 u_{t-1} - \hat{\theta}_2 u_{t-2}$:
  — pour l'horizon 1 : ${}_te_{t+1} = x_{t+1} -_t \hat{x}_{t+1} = u_{t+1}$
  — pour l'horizon 2 : ${}_te_{t+2} = x_{t+2} -_t \hat{x}_{t+2} = u_{t+2} - \theta_1 u_{t+1}$
  — pour les horizons $h \geq 3$: ${}_te_{t+3} = x_{t+h} -_t \hat{x}_{t+h} = u_{t+h} - \theta_1 u_{t+h-1} - \theta_1 u_{t+h-2}$

- Cas d'un ARMA(1,2). Avec $x_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \hat{u}_t - \hat{\theta}_1 \hat{u}_{t-1} - \hat{\theta}_2 \hat{u}_{t-2}$:
  — pour l'horizon 1 : ${}_te_{t+1} = x_{t+1} -_t \hat{x}_{t+1} = u_{t+1}$
  — pour l'horizon 2 : ${}_te_{t+2} = x_{t+2} -_t \hat{x}_{t+2} = u_{t+2} + (\hat{\phi}_1 - \hat{\theta}_2) u_{t+1}$
  — pour les horizons $h \geq 3$: ${}_te_{t+h} = x_{t+h} -_t \hat{x}_{t+h} = \hat{\phi}_1\, {}_te_{t+h-1} + u_{t+h} - \hat{\theta}_1 u_{t+h-1} - \hat{\theta}_2 u_{t+h-2}$

Vous pouvez constater que la variance de ces erreurs de prévision augmente avec l'horizon h. Comme on suppose que les coefficients estimés sont égaux aux vrais coefficients qui sont eux-mêmes des constantes (inconnues initialement), les termes $\hat{\phi}_i, i = 1\ldots, p$ et/ou $\hat{\theta}_i, i = 1\ldots, p$ sont aussi considérés comme étant des constantes. En conséquence, ayant estimé la variance du bruit blanc u, on peut estimer la variance des erreurs ${}_te_{t+h}$, soit $s^2_{{}_te_{t+h}}$ cette estimation.

Pour finir, comme $x_{t+h} =_t \hat{x}_{t+h} +_t e_{t+h}$, *i.e.* la réalisation future de la variable s'écarte de la valeur prévue en t d'un aléa d'espérance nulle dont on connaît maintenant la variance, il ne reste plus qu'à faire une hypothèse de distribution sur les résidus pour pouvoir construire un intervalle de confiance autour de la prévision. Si on pose que le bruit blanc est gaussien, alors pour un seuil de risque α quelconque, :

$$Pr\left[x_{t+h} \in \left[{}_t\hat{x}_{t+h} \pm z_{\alpha/2} \times s_{{}_te_{t+h}}\right] = 1 - \alpha \right. \tag{8}$$

où $z_{\alpha/2}$ est le fractile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\mathcal{N}(0, 1)$.

3 Les commandes de calcul des prévisions

3.1 sous SAS

On mobilise la commande `forecast` de la proc `arima`. Elle arrive naturellement après les commandes `identify` et surtout `estimate`. Les principales options sont les suivantes :

- `alpha = 0< nombre <1` : précise un niveau de risque, α, réclamant la construction d'intervalles de confiance au seuil 1-α. Par défaut $\alpha = 5\%$.
- `back = nombre entier` : permet de définir un rang à partir duquel le calcul des prévisions itérées va commencer. Par exemple, si on a 100 observations, `back=10` force le calcul des prévisions à partir de la $90^{\text{ème}}$ observation. Par défaut `back=0` : les prévisions démarrent après la dernière observation de la table.
- `lead = nombre entier` : donne le nombre de prévisions à construire. Par défaut, `lead=24`

Par défaut seul le graphique des prévisions itérées est construit. Pour obtenir celui des observations avec les valeurs calculées sur la période d'estimation (*i.e* des prévisions in-sample à l'horizon 1) et les prévisions itérées, il faut utiliser l'option `forecast(forecast)` ou `forecast(all)` dans la ligne d'appel de la procédure.

3.2 sous R

Si `"model"` est le nom d'une équation estimée par la commande arima appliquée à un objet de type *time series* selon :
`model <- arima(...)`,
alors il suffit d'utiliser la commande `forecast du package "forecast"` dans laquelle on va indiquer le nom du modèle à utiliser et le nombre de prévisions à construire. Par exemple, ici on fera :

```
prev <- forecast(model, h=30, level=0.90)
plot(prev)
```

pour obtenir le graphe des observations et des prévisions avec indication des intervalles de confiance à 90%. Notez qu'il est possible avec l'option `fan=T` d'obtenir un *fan chart*, *i.e.* une représentation colorée d'intervalles de confiance associés à plusieurs seuils de risque.