Méthodes de lissage exponentiel

- Typiquement dans un modèle de régression, on dispose des observations $(y_t, \mathbf{x}_t)$, $t = 1, \ldots, n$, avec n la taille de l'échantillon.
- On a alors formulé un modèle linéaire de la forme

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \epsilon_t.$$

- Pour faire des prévisions à l'instant t_0, on avait cependant besoin de $\mathbf{x}_{t_0}$.

Données mesurées dans le temps: séries chronologiques

- Il y a des situations où l'on dispose que d'une série de données.
- Lorsque les données sont mesurées dans le temps, il est anticipé que ces données soient dépendantes.
- L'hypothèse de l'échantillon aléatoire ($X_1, \ldots, X_n$ iid) devient alors douteuse.
- Lorsque les données sont mesurées dans le temps, on dira que $X_1, \ldots, X_n$ est une série chronologique.
- Formellement, une série chronologique est une réalisation finie d'un processus stochastique.

Calcul des prévisions

- Que faire si l'on dispose que d'une série de données et que l'on désire faire des prévisions?
- Une stratégie consiste à utiliser le passé de la variable, ou l'historique.
- En utilisant le passé de la variable, en expliquant la dépendance dans les données, on cherche à proposer des prévisions.

Méthode classique: lissage exponentiel

- Le lissage exponentiel est simple et intuitif; c'est l'ancêtre des méthodes plus modernes de séries chronologiques.
- Il demeure utile afin de motiver les nouveaux modèles, avec les outils vus jusqu'à maintenant.
- Considérons $z_1, \dots, z_n$ une série chronologique, réalisation de $\{Z_t, t \in \mathbb{Z}\}$.
- Un modèle possible est d'expliquer Z_t par le temps lui-même:

  $$Z_t = f(t, \beta) + \epsilon_t,$$

  où $f(t, \beta)$ est une fonction connue, fonction du temps, et les erreurs $\{\epsilon_t\}$ sont non-corrélées.

Exemples

- Dans le cas non-saisonnier, on pourrait considérer:

$$
\begin{aligned}
Z_t &= \beta + \epsilon_t, \text{ modèle avec moyenne constante;} \\
Z_t &= \beta_0 + \beta_1 t + \epsilon_t, \text{ modèle avec tendance linéaire;}
\end{aligned}
$$

- Cas saisonnier:

$$
\begin{aligned}
Z_t &= \beta_0 + \beta_1 \sin(2\pi t/12) + \beta_2 \cos(2\pi t/12) + \epsilon_t, \\
&\quad \text{c'est un exemple de modèle sinusoidal;} \\
Z_t &= \sum_{j=1}^{12} \beta_j I\{t \in \text{période } i\} + \epsilon_t, \text{ modèle avec indicatrices,}
\end{aligned}
$$

avec $I\{t \in \text{période } i\}$ valant un si t est dans la période i, $i = 1, \ldots, 12$.

Méthodes d'estimation

- Une première méthode d'estimation pourrait être les moindres carrés ordinaires:

$$\min \sum_{t=1}^{n} \{z_t - f(t, \beta)\}^2$$

- On verra cependant qu'une méthode plus naturelle consistera à utiliser les moindres carrés pondérés:

$$\min \sum_{t=1}^{n} w_t \{z_t - f(t, \beta)\}^2$$

- Les w_t sont des poids, qui sont choisis tels qu'ils décroissent de manière exponentielle:

$$w_t = w^{n-t}, \ t = 1, \ldots, n.$$

Choix du coefficient w

- Le coefficient w doit être choisi par l'utilisateur.
- Il aura un impact direct sur l'importance des observations récentes relativement aux données passées. Par exemple $w = 0.9$.
- Ce cadre des moindres carrés pondérés avec ce choix de poids mène à la méthode générale du lissage exponentiel.

Modèle avec moyenne constante

- Considèrons le modèle $Z_t = \beta + \epsilon_t$. Sous forme matricielle on aura alors:
$$\mathbf{Z} = \beta \mathbf{1} + \boldsymbol{\epsilon},$$
avec $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ et $\mathbf{1} = (1, \ldots, 1)^\top$.
- La variable à l'horizon l est alors:
$$Z_{n+l} = \beta + \epsilon_{n+l}.$$
- La prévision pour Z_{n+l} est alors:
$$Z_n(l) = \beta,$$
en autant que β soit connu.

Estimation de β dans le modèle avec moyenne constante

- Si on utilise les moindres carrés ordinaires, on trouve:

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}, \\
&= (\mathbf{1}^\top \mathbf{1})^{-1} \mathbf{1}^\top \mathbf{Z}, \\
&= n^{-1} \sum_{t=1}^{n} Z_t, \\
&= \bar{Z}_n.
\end{aligned}
$$

- On note que $\bar{Z}_n = \sum_{t=1}^{n} \{(1/n) Z_t\}$.
- L'utilisation des moindres carrés entraîne que chaque observation a un poids de $1/n$.

Modification des poids

- Les données z_t sont observées dans le temps.
- Intuitivement, il semble raisonnable d'attribuer plus de poids aux observations récentes, et moins aux observations passées.
- Une façon: pondérer de façon que les poids décroissent géométriquement dans le temps.
- Cette idée amène la prévision suivante:

$$\hat{z}_n(l) = c\left(z_n + wz_{n-1} + w^2 z_{n-2} + \ldots + w^{n-1} z_1\right),$$

avec $w \in (0, 1)$ et $c \sum_{t=0}^{n-1} w^t = 1$.

Justification théorique

- En fait, si on considère le problème des moindres carrés pondérés suivant:

$$\min_{\beta} \sum_{j=0}^{n-1} w^j \{z_{n-j} - \beta\}^2$$

- Les moindres carrés pondérés donnent:

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^\top \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{-1} \mathbf{y}, \\
&= \frac{\sum_{j=0}^{n-1} w^j z_{n-j}}{\sum_{j=0}^{n-1} w^j}, \\
&= \frac{1-w}{1-w^n} \sum_{j=0}^{n-1} w^j z_{n-j},
\end{aligned}$$

avec $\mathbf{X} = \mathbf{1} = (1, \ldots, 1)^\top$, $\mathbf{y} = (z_n, \ldots, z_1)^\top$ et $\mathbf{W} = \mathrm{diag}(1, 1/w, \ldots, 1/w^{n-1})$.

Facteur d'amortissement

- Òn rappelle que $c\sum_{t=0}^{n-1} w^t = 1$.
- Utilisant les résultats sur les séries géométriques:

$$\sum_{t=0}^{n-1} w^t = \frac{1-w^n}{1-w},$$

- Puisque $w \in (0,1)$, ceci implique:

$$c = \frac{1-w}{1-w^n} \to 1-w,$$

  lorsque $n \to \infty$.
- Typiquement $w \in (0.7, 0.95)$.

Étude en fonction de w

- Avec $c = \frac{1-w}{1-w^n}$, on remarque que la prévision devient:

$$\hat{z}_n(l) = \frac{1-w}{1-w^n}\left(z_n + wz_{n-1} + w^2 z_{n-2} + \ldots + w^{n-1}z_1\right),$$

- Le poids associé à chaque donnée est donc $\frac{1-w}{1-w^n}w^j$. On aura:

$$\begin{aligned} \lim_{w\to 1}\frac{1-w}{1-w^n} &= \frac{0}{0}; \\ \lim_{w\to 1}\frac{-1}{-nw^{n-1}} &= \frac{1}{n}. \end{aligned}$$

- Si w est proche de un, tous les poids sont grands ou plutôt de même importance (et proches de $1/n$).
- Les prévisions seront lisses et on parlera d'un lissage fort.
- Si $w << 1$, les prévisions vont reposer sur les dernières données; ce sera moins lisse.

Astuce décisive: connaissance du passé infini

- La prévision devrait dépendre sur les données $z_1, \ldots, z_n$:

$$\hat{z}_n(l) = \frac{1-w}{1-w^n}\left(z_n + wz_{n-1} + w^2 z_{n-2} + \ldots + w^{n-1} z_1\right),$$

- Cependant, compte tenu que $w^n \to 0$ rapidement, et compte tenu d'astuces théoriques, il est souvent commode de supposer que l'on dispose de tout l'historique:

$$\hat{z}_n(l) = (1-w)\left(z_n + wz_{n-1} + w^2 z_{n-2} + \ldots\right),$$

où l'on s'est permis de laisser tendre $n \to \infty$ et donc $\frac{1-w}{1-w^n} \to 1-w$, lorsque $n \to \infty$.

Premier lissage des observations

- La prévision:

$$\hat{z}_n(l) = (1 - w)\left(z_n + wz_{n-1} + w^2 z_{n-2} + \dots\right),$$

  est appelée *premier lissage des observations*.
- Le facteur $\lambda = 1 - w$ est appelée *constante de lissage*.
- On a mentionné que $w \in (0.7, 0.95)$, donc typiquement $\lambda \in (0.05, 0.3)$. Ce sont des règles empiriques.
- Le premier lissage est parfois noté S_n:

$$S_n = \lambda \sum_{j=0}^{\infty} w^j z_{n-j}, \ \forall l \geq 1.$$

- On note que S_n est indépendant de l'horizon l

Mise à jour des prévisions

- Sujet qui historiquement a assuré le succès de la méthode.
- Repose sur une utilisation astucieuse de la connaissance du passé infini (pas immédiatement évident à réaliser).
- Ceci repose sur l'argumentation suivante:

$$
\begin{aligned}
\hat{z}_n(l) = S_n &= \lambda \sum_{j=0}^{\infty} w^j z_{n-j} = \lambda z_n + \lambda \sum_{j=1}^{\infty} w^j z_{n-j}, \\
&= \lambda z_n + \lambda w \sum_{j=1}^{\infty} w^{j-1} z_{n-j} = \lambda z_n + \lambda w \sum_{j=0}^{\infty} w^j z_{n-j-1}, \\
&= \lambda z_n + w \left\{ \lambda \sum_{j=0}^{\infty} w^j z_{(n-1)-j} \right\}, \\
&= \lambda z_n + w S_{n-1}.
\end{aligned}
$$

Mise à jour des prévisions; formules équivalentes

- On vient de voir:

$$\hat{z}_n(l) = S_n = \lambda z_n + w S_{n-1}.$$

- On peut écrire également:

$$\hat{z}_n(1) = \lambda z_n + w \hat{z}_{n-1}(1).$$

- On remarque que $\hat{z}_n(1)$ est la prévision d'horizon un.
- Une autre écriture est:

$$\begin{aligned} S_n &= (1-w) z_n + (w - 1 + 1) S_{n-1}, \\ &= S_{n-1} + (1-w)(z_n - S_{n-1}). \end{aligned}$$

Modèle à correction d'erreurs

- La formule:

$$S_n = S_{n-1} + (1 - w)(z_n - S_{n-1}).$$

peut s'écrire alternativement comme:

$$\hat{z}_n(1) = \hat{z}_{n-1}(1) + (1 - w)(z_n - \hat{z}_{n-1}(1)).$$

- L'expression précédente est utile car elle indique comment les prévisions sont mises à jour quand une nouvelle observation devient disponible.
- C'est un exemple de modèle à correction d'erreurs.

Mise en oeuvre en pratique

- On postule l'utilisation de $S_n = \lambda \sum_{j\geq 0} w^j z_{n-j}$.
- On exploite les récursions $S_n = \lambda z_n + wS_{n-1}$.
- Les récursions donnent:

$$
\begin{aligned}
S_n &= \lambda z_n + w(\lambda z_{n-1} + wS_{n-2}), \\
&= \lambda z_n + \lambda w z_{n-1} + w^2 S_{n-2}, \\
&= \ldots, \\
&= \lambda z_n + \lambda w z_{n-1} + \lambda w^2 z_{n-2} + \ldots + w^n S_0.
\end{aligned}
$$

- On aura besoin de λ et de S_0: la prévision S_n devient calculable.

Choix de S_0

- La formule est:

$$S_n = \lambda z_n + \lambda w z_{n-1} + \lambda w^2 z_{n-2} + \ldots + w^n S_0.$$

- Si n est grand et w petit alors l'influence de S_0 devrait être faible.
- Si w est proche de un, on peut prendre $S_0 = \bar{z}_n$.
- Si w est loin de un, on pourrait prendre $S_0 = z_1$.

Choix de λ

- On considère les erreurs de prévision d'horizon un suivantes:

$$\begin{aligned} e_t(1) &= z_{t+1} - S_t, \\ &= z_{t+1} - \hat{z}_t(1),\ t = 1, \ldots, n, \end{aligned}$$

- On considère la somme des carrés suivante:

$$\text{RSS}(\lambda) = \sum_{t=0}^{n-1} e_t^2(1).$$

- On minimise en λ:

$$\min_{\lambda} \text{RSS}(\lambda)$$

- En pratique on peut choisir une palette de valeurs de $\lambda \in \{0.05, 0.06, \ldots, 0.30\}$.

Diagnostiquer un lissage exponentiel

- Les 'résidus' dans ce cas-ci sont les erreurs de prévision d'horizon un:

$$e_t(1) = z_{t+1} - S_t,\ t = 0, \ldots, n-1.$$

- Deux aspects sont souvent cernés:
  1. Absence de corrélation dans les erreurs de prévision. On fait un test de bruit blanc sur les $e_t(1)$:

$$r(k) = \frac{\sum_{t=k}^{n-1} \{e_t(1) - \bar{e}\}\{e_{t-k}(1) - \bar{e}\}}{\sum_{t=0}^{n-1} \{e_t(1) - \bar{e}\}^2}$$

  2. Prévisions non-biaisées. On teste formellement $H_0 : \mu_e = 0$, avec μ_e l'espérance des erreurs de prévision un.

Diagnostiquer un lissage exponentiel (suite)

- Pour tester $\mu = 0$ sur un échantillon $X_1, \ldots, X_n$ la statistique est $n^{1/2}\bar{x}/s$, avec $\bar{x}$ et s^2 la moyenne et la variance échantillonnales, respectivement.
- Les 'observations' sont les $e_t(1) = z_{t+1} - S_t$, $t = 0, \ldots, n-1$.
- Pour tester formellement $H_0 : \mu_e = 0$, on calcule la statistique:

$$n^{1/2}\bar{e}/s_e,$$

avec $\bar{e} = n^{-1}\sum_{t=0}^{n-1} e_t$ et $s_e^2 = n^{-1}\sum_{t=0}^{n-1}\{e_t(1) - \bar{e}\}^2$.
- On rejette H_0 au niveau 5% si:

$$|n^{1/2}\bar{e}/s_e| > 1.96.$$

Intervalle de prévision

- On remarque que:

$$\begin{aligned} \mathrm{var}(S_n) &= \mathrm{var}(\lambda \sum_{j\geq 0} w^j z_{n-j}) = \lambda^2\sigma^2 \sum_{j\geq 0} w^{2j}, \\ &= \sigma^2\lambda^2 \frac{1}{1-w^2}. \end{aligned}$$

- Ainsi l'erreur quadratique moyenne de prévision $E\{(Z_{n+1} - S_n)^2\}$ devient:

$$E\{(Z_{n+1} - S_n)^2\} = \sigma^2 + \sigma^2\lambda^2 \frac{1}{1-w^2} = \sigma^2 \frac{2\lambda}{1-w^2}$$

- Il peut être montré qu'un intervalle de prévision est:

$$S_n \pm q_{1-\alpha/2} s_e,$$

avec $s_e^2 = n^{-1} \sum_{t=0}^{n-1} \{e_t(1) - \bar{e}\}^2$.