Séries chronologiques univariées (STT-6615)

Chapitre 2

Construction des modèles ARIMA

Les étapes principales pour construire un modèle ARIMA

- Représentation graphique des données;
- Transformation des données afin de rendre la série stationnaire;
- Identification préliminaire des ordres autorégressifs et moyennes mobiles suivant les caractéristiques générales des modèles ARMA;

Les étapes principales pour un construire un modèle ARIMA (suite)

- Estimation des paramètres du modèle;
- Diagnostiques;
- Choix du modèle final.

Étape d'identification

- Est-ce que la variance semble constante?
  - On veut s'assurer que le terme de variance dans le modèle est constant, par exemple comme fonction du temps.
  - Transformations populaires:
    - Transformation *racine*, transformation *inverse*, transformation *logarithmique*.
    - Préférablement (mais de manière optionnelle dans le cours), méthodologie de Box-Cox pour trouver la transformation.
  - On veut stabiliser la variance, et en plus, se rapprocher de la normalité des erreurs.

Transformation de Box-Cox

- La transformation a la forme:

$$y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log(y), & \lambda = 0. \end{cases}$$

- Le choix de λ se fait souvent en effectuant un graphique de la vraisemblance en fonction de λ.

Remarque

- L'application peut suggérer une transformation.
- **Exemple:** séries chronologiques financières.
- On pourrait disposer de:

$$X_t = (1 + P_t) X_{t-1}$$

- Ici X_t pourrait représenter la valeur d'un investissement au temps t et P_t est le changement en pourcentage pour la période allant de $t-1$ à t.

Remarque (suite)

- Prenant le logarithme:

$$\log(X_t) = \log(1 + P_t) + \log(X_{t-1}).$$

- Ce qui implique:

$$\log(X_t) - \log(X_{t-1}) = \log(1 + P_t),$$

$$\nabla \log(X_t) = \log(1 + P_t),$$

$$\approx P_t,$$

- sous l'hypothèse que le changement en pourcentage reste relativement petit en magnitude.

Identification (suite)

- 2. Choix du niveau de différentiation. Ici on veut une série stationnaire. En particulier la moyenne ne doit pas dépendre du temps.
- Si la série est stationnaire: Si $\{X_t\}$ est ARMA(p,q), alors les autocorrélations $\rho(k)$ sont en nombre infini et décroissent vers 0 *plutôt rapidement*.
- Un estimateur de $\rho(k)$ est donné par $\hat{\rho}(k)$.

Identification (suite)

- De plus, on devrait avoir que:

$$\hat{\rho}(k) \xrightarrow{P} \rho(k)$$

- Cependant, la série $\{X_t\}$ est non-stationnaire, que se passe-t-il? On sait que $cov(X_t, X_{t-k})$ va dépendre de k, mais aussi de t. Ainsi les autocorrélations $\rho(k)$ ne sont pas définies. Mais que sera la comportement des statistiques $\hat{\rho}(k)$?

Théorème

- Soit une série chronologique $X_1, X_2, \ldots, X_n$ générée d'un processus ARIMA(*p*,*d*,*q*), où $d \geq 1$; on présume que le terme d'erreur $\{w_t\}$ est un bruit blanc Gaussien.
- Alors pour chaque délai *k* fixé, on a que:

$$\hat{\rho}(k) \xrightarrow{P} 1, \quad 1 \leq k \leq K < n$$

Cas stationnaire versus cas non-stationnaire, comportement des $\hat{\rho}(k)$, ***k*** fixé, comme fonction de ***n***.

- Cas stationnaire:

$$\hat{\rho}(k) \xrightarrow{P} \rho(k)$$

- Cas non-stationnaire:

$$\hat{\rho}(k) \xrightarrow{P} 1, \quad 1 \le k \le K < n$$

Autres éléments d'information

- **Cas stationnaire:** Les $\hat{\rho}(k)$ comme fonction de *k* décroissent à 0 de façon exponentielle (décroissance vers 0 très rapide).
- **Cas non-saisonnier:** à partir de *k = 20*, toutes les autocorrélations devraient être très près de 0.
- **Cas non-stationnaire:** Décroissance des $\hat{\rho}(k)$ de manière linéaire vers 0? Décroissante très lente? Est-ce que les $\hat{\rho}(k)$ sont toutes de même signes? Ce sont tous des indices d'un problème de stationnarité.

Remarque concernant la sur différenciation

- Il faut éviter de différencier un trop grand nombre de fois.
- Exemple: Supposons que le processus $\{X_t\}$ est un bruit blanc. On aura que la série différenciée est un MA(1), introduisant ainsi de la dépendance inutile.
- Si on ajuste un MA(1) typiquement on aura une valeur de $\hat{\theta}$ proche de un, suggérant un problème d'inversibilité.

Identification: choix de *p* et de *q*

- Ayant identifié *d*, on a maintenant comme modèle $\{W_t = \nabla^d X_t\}$.
- Important éléments d'information:
  - Pour un *autorégressif*, toutes les autocorrélations partielles s'annulent après un certain délai.
  - Pour un *moyenne-mobile*, toutes les autocorrélations s'annulent après un certain délai.
- Que fait-on si les $\rho(k)$ et les ϕ_{kk} semblent tous les deux en nombre infini?

Modélisation des résidus

- Supposons que le processus est:

$$\phi(B)(1-B)^d X_t = \theta(B) v_t, \quad \forall t.$$

- Cependant, le processus $\{v_t\}$ n'est pas un bruit blanc mais un ARMA$(\tilde{p},\tilde{q})$. Donc:

$$\tilde{\phi}(B) v_t = \tilde{\theta}(B) w_t, \quad \forall t.$$

$$v_t = \tilde{\phi}^{-1}(B)\tilde{\theta}(B) w_t$$

Modélisation des résidus (suite)

- Quelques manipulations donnent:

$$\phi(B)(1-B)^d X_t = \theta(B)v_t, \quad \forall t.$$

$$\phi(B)(1-B)^d X_t = \theta(B)\tilde{\phi}^{-1}(B)\tilde{\theta}(B)w_t,$$

$$\tilde{\phi}(B)\phi(B)(1-B)^d X_t = \theta(B)\tilde{\theta}(B)w_t$$

$$\phi^*(B)(1-B)^d X_t = \theta^*(B)w_t,$$

- Ce modèle n'est rien d'autre qu'un ARIMA$\left(p^*, d, q^*\right)$

Modélisation des résidus: exemple

- Supposons que vos disposez d'une série. Vous la différenciez une fois pour la rendre stationnaire. Vous disposez à ce stade d'un ARIMA(*p,1,q*).
- Vous trouvez que $\hat{\gamma}(1)$ suggère une composante MA. Vous modéliser un ARIMA(*0,1,1*).
- Un examen des résidus suggère une composante AR*(1)*. Vous tentez finalement un ARIMA*(1,1,1)* qui devrait faire l'affaire.
- **En résumé: ce sont les résidus qui permettent de construire des modèles ARMA en pratique.**