Séries Temporelles Avancées

Laurent Ferrara

M2 EIPMC // Univ. Paris Ouest / Sep. 2013

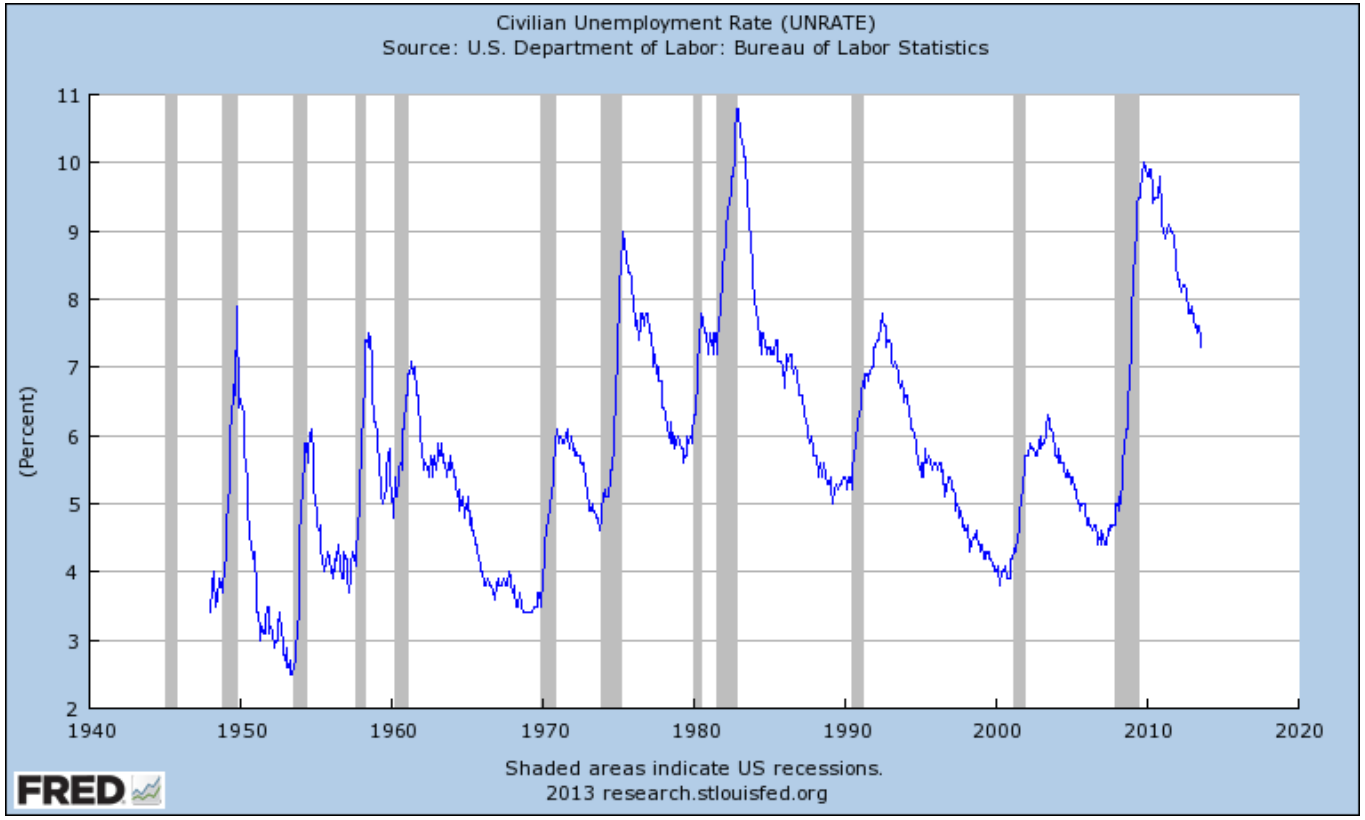
Plan

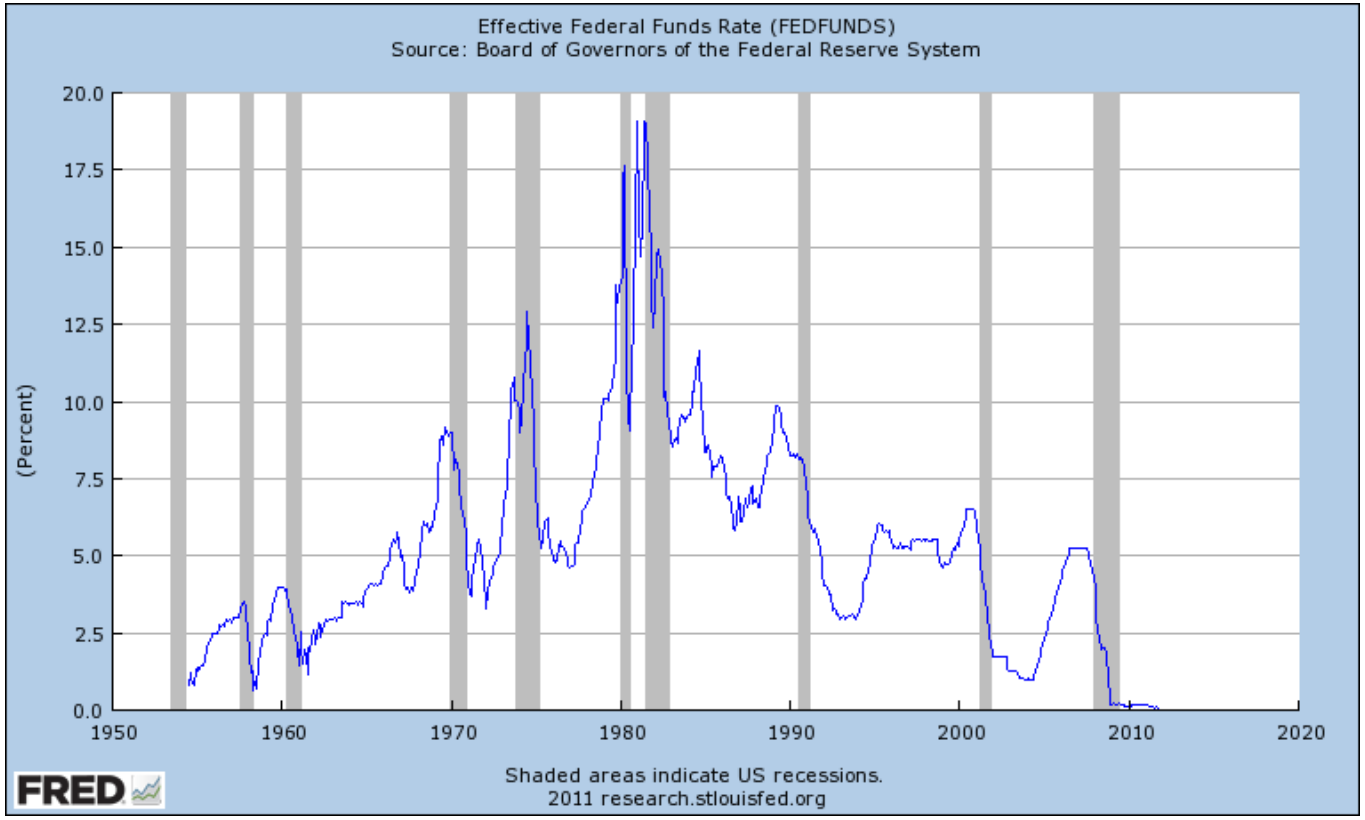
Base

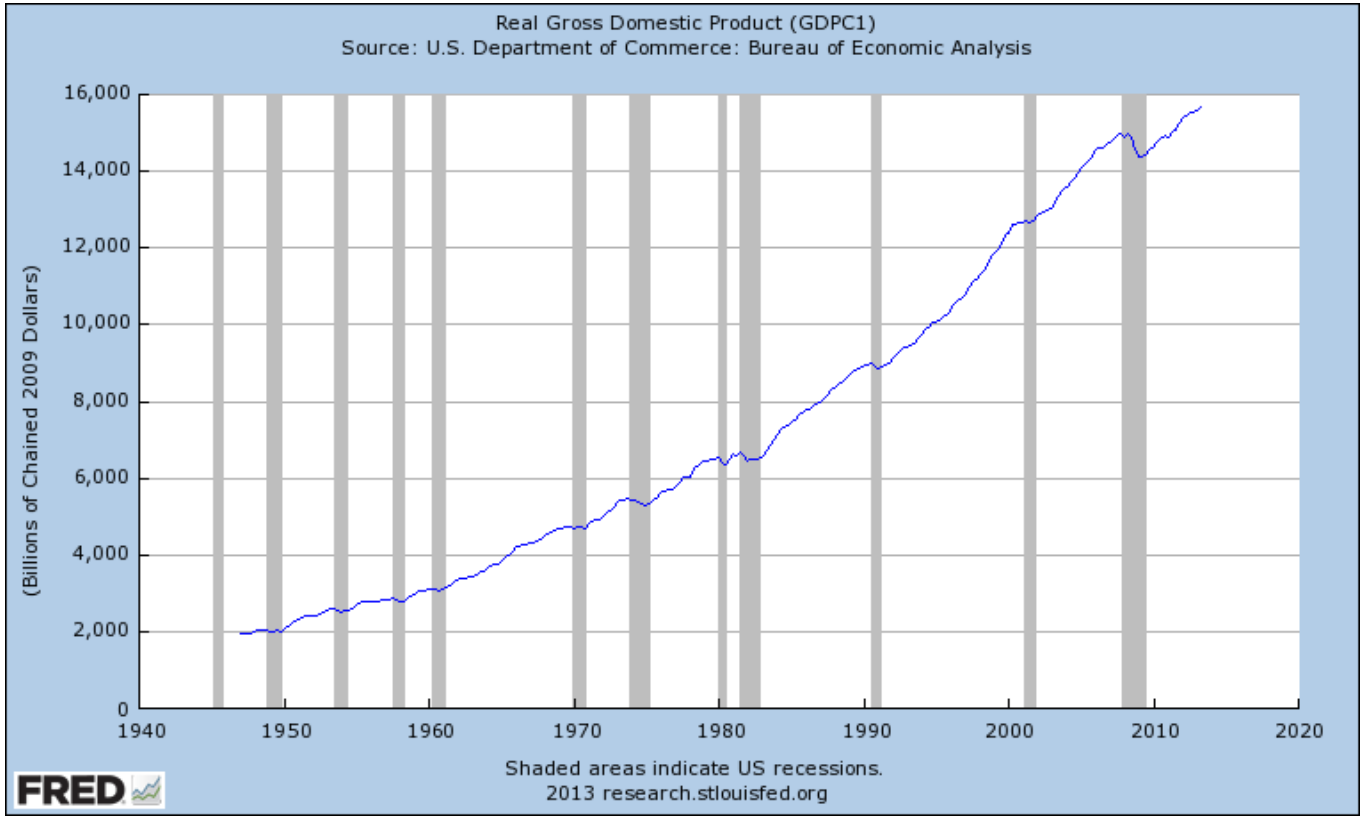
Comment identifier un processus candidat ?

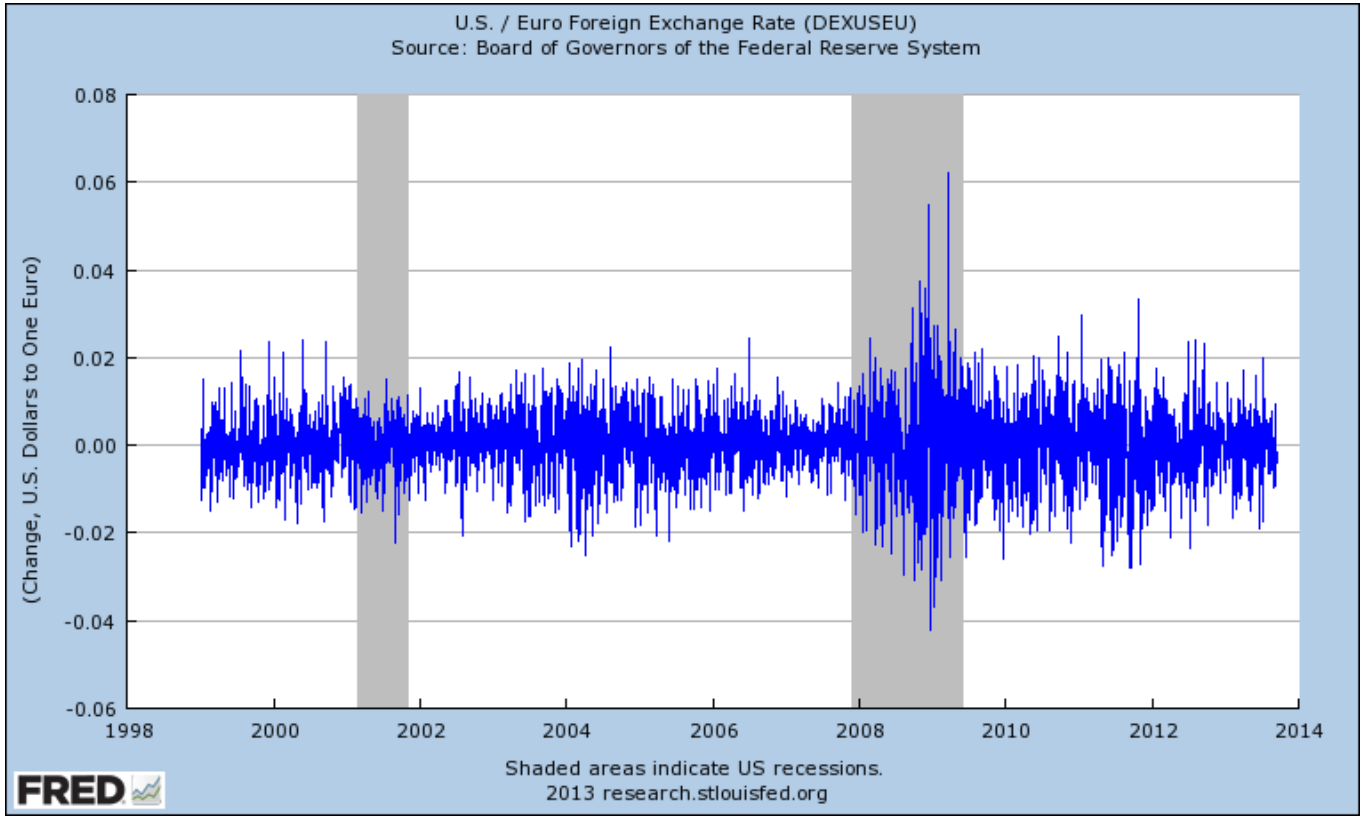
Prév. linéaire

Faits stylisés des séries financières


Civilian Unemployment Rate (UNRATE)
Source: U.S. Department of Labor: Bureau of Labor Statistics
(Percent)
11
10
9
8
7
6
5
4
3
2
1940
1950
1960
1970
1980
1990
2000
2010
2020
Shaded areas indicate US recessions.
2013 research.stlouisfed.org
FRED


Effective Federal Funds Rate (FEDFUNDS)
Source: Board of Governors of the Federal Reserve System
(Percent)
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
1950
1960
1970
1980
1990
2000
2010
2020
FRED
Shaded areas indicate US recessions.
2011 research.stlouisfed.org


Real Gross Domestic Product (GDPC1)
Source: U.S. Department of Commerce: Bureau of Economic Analysis
(Billions of Chained 2009 Dollars)
16,000
14,000
12,000
10,000
8,000
6,000
4,000
2,000
0
1940
1950
1960
1970
1980
1990
2000
2010
2020
Shaded areas indicate US recessions.
2013 research.stlouisfed.org
FRED


U.S. / Euro Foreign Exchange Rate (DEXUSEU)
Source: Board of Governors of the Federal Reserve System
(Change, U.S. Dollars to One Euro)
0.08
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
1998
2000
2002
2004
2006
2008
2010
2012
2014
Shaded areas indicate US recessions.
2013 research.stlouisfed.org
FRED

Definition

Un processus $(X_t)_{t \in Z}$ est une famille de variables aléatoires à valeurs réelles indéxée par $t \in Z$.

Pour une valeur de ω fixée dans Ω, la fonction qui associe à chaque date t la réalisation $X_t(\omega)$ est la trajectoire du processus au point ω.
On note $(x_1, \ldots, x_t, \ldots x_T)$ la trajectoire observée pour $t = 1, \ldots, T$.

Si le processus a été spécifié, estimé et validé, on peut alors l'utiliser pour effectuer une prévision.
On construit alors l'estimateur $\hat{X}_T(h)$ qui est le prédicteur de la variable aléatoire X_{T+h}.
Comme tout estimateur, ce prédicteur est à son tour une variable aléatoire, en tant que fonction mesurable de v.a.
Ainsi, $\hat{X}_T(h)$ possède une loi de distribution, qu'il conviendra de spécifier dans la mesure du possible.

Comment se présente l'information dans une trajectoire ?

Ce qui caractérise une trajectoire $(x_1, \ldots, x_T)$ issue d'un processus est la non indépendance des v.a. $(X_1, \ldots, X_T)$.
Outil standard de mesure de la dépendance: coefficient de corrélation linéaire:

Definition

Soit $(X_t)_{t\in Z}$ un processus du second ordre (*i.e.* : $E(X_t^2) < \infty$).
(i) La fonction moyenne, notée $m(.)$, est l'espérance non conditionnelle du processus, *i.e.*: $m(t) = E(X_t)$.
(ii) La fonction d'autocovariance au retard k, notée $\gamma(k)$, est définie, pour tout $t \in Z$ et $k \in Z$, :

$$\gamma(k) = cov(X_t, X_{t+k}) = E\left[(X_t - E(X_t))(X_{t+k} - E(X_{t+k}))\right]. \quad (1)$$

(iii) La fonction ACF au retard k, notée $\rho(k)$, est définie de la manière suivante, pour tout $t \in Z$ et $k \in Z$, :

$$\rho(k) = \frac{\gamma(k)}{\sigma_{X_t}\sigma_{X_{t+k}}}, \quad (2)$$

où σ_{X_t} est l'écart type du processus à t, tel que : $\sigma_{X_t} = \sqrt{\gamma(0)}$.

Autre mesure : la PACF

La PACF au retard k, notée $r(k)$, est définie pour tout $k \in Z$, de la manière suivante :

$$r(k) = \frac{cov(X_t - X^*, X_{t+k} - X^*_{t+k})}{var(X_t - X^*)^{1/2} var(X_{t+k} - X^*_{t+k})^{1/2}}, \qquad (3)$$

où, pour tout t, X^*_t est la régression affine de X_t sur $X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}$ et X^*_{t+k} est la régression affine de X_{t+k} sur $X_{t+k-1}, X_{t+k-2}, \ldots, X_{t+1}$.

Proposition 1 *Le coefficient $r(k)$ défini par l'équation 3 est le coefficient de X_t dans la régression linéaire de X_{t+k} sur $1, X_t, X_{t+1}, \ldots, X_{t+k-1}$.*

En pratique à échantillon fini

Estimateur de la moyenne:

$$\bar{X}_T = T^{-1} \sum_{t=1}^{T} X_t$$

La fonction d'autocovariance d'un processus au retard k est estimée par la fonction d'autocovariance empirique, $\hat{\gamma}(.)$, définie, pour $0 \leq k < T$, par :

$$\hat{\gamma}(k) = \frac{1}{T} \sum_{t=1}^{T-k} (X_t - \bar{X}_T)(X_{t+k} - \bar{X}_T). \tag{4}$$

De même, l'ACF est estimée par l'ACF empirique, notée $\hat{\rho}(.)$ et définie, pour $0 \leq k < T$, par :

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\sigma_{X_t}} \hat{\sigma_{X_{t+k}}}}. \tag{5}$$

On remarque également que la matrice de corrélation estimée, $\hat{R} = [\hat{\rho}(i-j)]_{i,j=1,\ldots,T}$, est définie positive.

Comment identifier un processus candidat ?

- L'ACF fournit une mesure de la persistence ou de la **mémoire** du processus.
- A partir de cette information, nous allons chercher quel type de processus permet de reproduire cette persistence.
- Nous allons caractériser trois de types de mémoire : mémoire longue, mémoire courte et sans mémoire.

Absence de mémoire

Definition

Un processus d'ordre 2 $(\varepsilon_t)_{t\in Z}$ est un processus bruit blanc faible si :

(i) $\forall t,\ E(\varepsilon_t) = 0$

(ii) $\forall t, \forall s,\ E(\varepsilon_t \varepsilon_s) = \sigma_\varepsilon^2 \times I_{[t=s]}$

où $I(.)$ est la fonction indicatrice.

Bruit blanc fort: bruit blanc indépendant.

Mémoire courte

Definition

Un processus est dit à mémoire courte s'il possède une ACF, $\rho(k)$, telle que :

$$\rho(k) \leq Cr^{-k}, \rightarrow \infty, \tag{6}$$

où $C > 0$, $0 < r < 1$ et $k = 1, 2, \ldots$.

Exemples de processus mémoire courte

Exemple 1

Un processus moyenne-mobile d'ordre 1, de la forme suivante :

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

où ε_t est un processus bruit blanc faible, est un processus à mémoire courte. En général, pour des raisons d'inversibilité et d'indentifiabilité le paramètre θ est tel que : $|\theta| < 1$. Pour ce processus, on montre que $E(X_t) = 0$, et que $\rho(1) = \theta$ et $\rho(k) = 0$ si $k > 1$.

Exemples de processus mémoire courte

Exemple 2

Un processus autoregressif d'ordre 1, de la forme suivante :

$$X_t - \phi X_{t-1} = \varepsilon_t$$

où ε_t est un processus bruit blanc faible, est un processus à mémoire courte. En général, pour des raisons d'inversibilité et de stationnarité (voir ci-après), le paramètre ϕ est tel que : $|\phi| < 1$. Pour ce processus, on montre que $E(X_t) = 0$, et que $r(1) = \phi$ et $r(k) = 0$ si $k > 1$.

Mémoire longue

Si l'ACF est non nulle pour des retards élevés, en pratique de l'ordre de $k \geq 20$, on dit que le processus est fortement persistant. Plus formellement, on parle de processus à mémoire longue lorsque l'ACF du processus, $\rho(k)$, décroît comme une fonction puissance de k.

Definition

Un processus est dit à mémoire longue s'il possède une ACF, $\rho(k)$, t.q.:

$$\rho(k) \sim Ck^{-\alpha} \text{ quand } k \to \infty, \tag{7}$$

où $\sim$ représente l'équivalence asymptotique, où $C > 0$ est une constante et où $\alpha \in]0, 1[\in R$.

Alors: $\sum_{k=0}^{\infty} |\rho(k)| = \infty$. Les processus intégrés fractionnaires ARFIMA reproduisent ce fait stylisé.

Exemples de processus mémoire courte

Exemple 3

Le processus fractionnaire intégré introduit par Hosking (1980) et Granger et Joyeux (1981) de la forme suivante :

$$(I - B)^d X_t = \varepsilon_t$$

où B est l'opérateur retard tel que $B(X_t) = X_{t-1}$ et $B^k(X_t) = X_{t-k}$ et d est un réel fractionnaire tel que $0 < d < 1$, est un processus à mémoire longue.

Non indépendant, mais identiquement distribué ?

Definition

Un processus $(X_t)_{t\in Z}$ est dit fortement stationnaire si, $\forall t_1, \ldots, t_n \in Z$, $\forall k \in Z$ et $n = 1, 2, \ldots$, la loi du vecteur $(X_{t_1}, \ldots, X_{t_n})$ est identique à la loi du vecteur $(X_{t_1+k}, \ldots, X_{t_n+k})$, i.e. toutes les lois de dimension finie du processus sont identiques.

Definition

Un processus du second ordre $(X_t)_{t\in Z}$ est dit faiblement stationnaire si :

(i) la moyenne du processus est constante au cours du temps, *i.e.* : pour tout $t \in Z$, $E(X_t) = \mu$,

(ii) la covariance du processus est invariante au cours du temps, *i.e.* : pour tout $t \in Z$ et $k \in Z$, $\gamma(k)$ ne dépend que de k.

Un processus faiblement stationnaire est également appelé stationnaire au second ordre, stationnaire en covariance ou stationnaire. Si le processus est faiblement stationnaire, l'espérance de chaque variable est identique et on peut alors l'estimer par la moyenne empirique $\bar{X}_T$. Ainsi, on peut centrer tout processus stationnaire en lui retranchant sa moyenne empirique.

Remarque 1 *Un processus fortement stationnaire est faiblement stationnaire, l'inverse n'étant généralement pas vrai. Un contre-exemple est le processus Gaussien pour lequel les deux types de stationnarités sont équivalents.*

Prédicteur naturel d'un processus stationnaire

La stationnarité d'un processus permet ainsi d'estimer les moments non conditionnels de la v.a. X_{T+h} en utilisant les moments empiriques à partir du processus $(X_1, \ldots, X_T)$. Ainsi, on peut utiliser comme prédicteur naturel de X_{T+h} un estimateur de l'espérance non conditionnelle $E(X_{T+h})$, en particulier la moyenne empirique, *i.e.* : $\hat{X}_{T+h} = \bar{X}_T = T^{-1} \sum_{t=1}^{T} X_t$. Ainsi, la prévision est alors obtenue, pour tout h, par $\hat{x}_T(h) = T^{-1} \sum_{t=1}^{T} x_t$.

Comment caractériser la qualité d'un prédicteur ?

Un "'bon"' estimateur d'une valeur: sans biais et de variance minimale.
Dans le cas de la variable $\hat{X}_T(h)$, le prédicteur à la data T pour l'horizon h ($h > 0$) de X_{T+h}, on introduit la variable d'erreur de prévision à l'horizon h définie par :

$$e_{T+h} = X_{T+h} - \hat{X}_T(h) \tag{8}$$

En généralisant l'erreur de prévision au temps T précédente à l'ensemble des temps, on introduit le processus d'erreur de prévision $(e_{t+h})_{t \in Z}$ tel que : $e_{t+h} = X_{t+h} - \hat{X}_t(h)$, pour tout $t \in Z$ et tout $h > 0$.

Critères de mesure de qualité de prévision

$$ME = E(e_{T+h}) \tag{9}$$

$$MAE = E(|e_{T+h}|) \tag{10}$$

$$MSE = E(e_{T+h}^2) \tag{11}$$

$$RMSE = \sqrt{E(e_{T+h}^2)} \tag{12}$$

Critères de mesure de qualité de prévision

En pratique, on compare $(x_{T+1}, \ldots, x_{T+h})$ avec $(\hat{x}_{T+1}, \ldots, \hat{x}_{T+h})$:

$$MSE = \frac{1}{h} \sum_{k=1}^{h} (x_{T+k} - \hat{x}_{T+k})^2 \tag{13}$$

Prévision par processus linéaires

Definition

Un processus $(X_t)_{t \in Z}$ est un processus linéaire s'il admet une décomposition de la forme suivante, $\forall t \in Z$:

$$X_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i}, \tag{14}$$

où :
(i) les coefficients $(a_i)_i$ sont absolument sommables, i.e.:
$\sum_{i=-\infty}^{\infty} |a_i| < \infty$,
(ii) $(\varepsilon_t)_t$ est un processus bruit blanc fort.

Meilleur prédicteur

Si on observe une trajectoire $(x_1, \ldots, x_T)$ que l'on suppose engendrée par une processus linéaire $(X_t)_{t \in Z}$, on connait alors le meilleur prédicteur $\hat{X}_T(h)$, au sens de la plus faible erreur quadratique moyenne. On note I_T l'ensemble d'information apporté par les variables $(X_1, \ldots, X_T)$, qui est en terme probabiliste la σ-algèbre engendrée par les T v.a.. On note M_T le sous-espace vectoriel fermé engendré par les variables $(X_1, \ldots, X_T)$, muni du produit scalaire $\langle X_t, X_{t'} \rangle = E(X_t X_{t'})$. La norme issue du produit scalaire est la norme L^2, notée $\|.\|_{L^2}^2$.

Proposition 2 *Le prédicteur $\hat{X}_T(h)$ qui minimise l'erreur quadratique moyenne (MSE) est le prédicteur des moindres carrés définie par :*

$$\hat{X}_T(h) = E(X_{T+h} | I_T),$$

soit :

$$\hat{X}_T(h) = \arg \min_{Y \in M_T} \|X_{T+h} - Y\|_{L^2}^2 , \qquad (15)$$

Processus d'innovation

Definition

On définit le processus d'innovation $(\epsilon_t)_t$ d'un processus $(X_t)_{t\in Z}$ comme étant l'écart entre la variable X_t au temps t et sa projection sur l'espace vectoriel engendré par les variables jusqu'au temps (t-1), i.e.:

$$\epsilon_t = X_t - E(X_t|I_{t-1})$$

Le processus d'innovation d'un processus stationnaire est un bruit blanc et un processus bruit blanc est son propre processus d'innovation.

Distribution du prédicteur

Pour un processus linéaire, on a $E(e_{T+h}) = 0$, et de variance telle que :

$$E(e_{T+h}^2) = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} a_i^2, \tag{16}$$

avec $a_0 = 1$.
Sous l'hypothèse de connaissance de la loi du processus d'erreur de prévision, on peut calculer un intervalle de confiance pour la prévision au niveau de confiance $1 - \alpha$:

$$X_{T+h} \in [\hat{X}_T(h) \pm t_{1-\alpha/2}\sigma_\varepsilon \sqrt{\sum_{i=0}^{h-1} a_i^2}], \tag{17}$$

où $t_{1-\alpha/2}$ est le quantile de la loi d'ordre $1 - \alpha$.

Distribution du prédicteur

Si on ne connaît pas la loi de distribution:
Méthodes de rééchantillonage de type Bootstrap
Voir par exemple Bec, Bouadbdallah et Ferrara (2013, WP BdF)
dans un cadre plus large (non-linéaire)

Quel horizon de prévision ? Court, moyen ou long terme ?

Si le processus est stationnaire, on montre que la loi de distribution conditionnelle converge vers la loi de distribution non conditionnelle lorsque l'horizon tend vers l'infini, ie:

$$L(X_{T+h}|I_T) \overset{h\to\infty}{\to} L(X_1) \tag{18}$$

Exemple

Soit un processus AR(1) stationnaire de moyenne nulle:

$$X_t - \phi X_{t-1} = \varepsilon_t$$

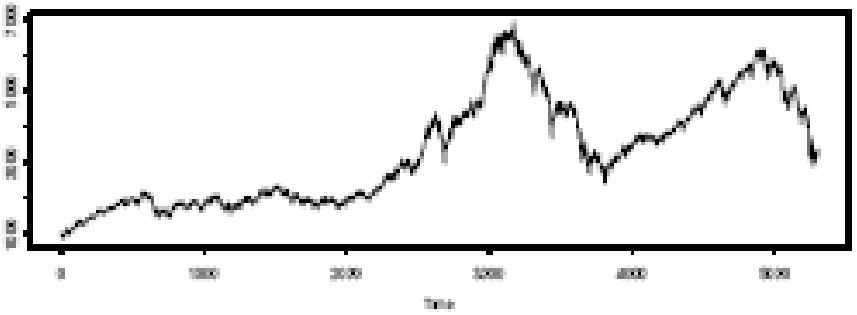
où ε_t est un processus bruit blanc faible et ϕ est tel que : $|\phi| < 1$.
Pour tout t, le prédicteur à l'horizon $h = 1$, noté $\hat{X}_t(1)$ est donné par

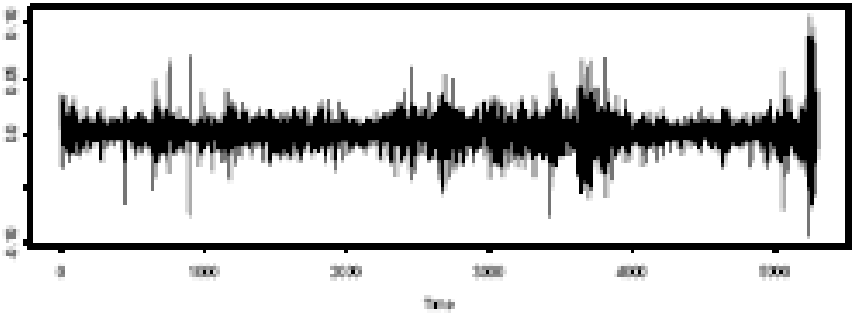
$$\hat{X}_t(1) = E(X_{t+1}|I_t) = \phi E(X_t|I_t) + E(\varepsilon_{t+1}|I_t) = \phi X_t.$$

De même, pour tout $h > 0$, on montre que :

$$\hat{X}_t(h) = \phi^h X_t.$$

Ainsi, lorsque $h \to \infty$, $\hat{X}_t(h)$ converge vers son espérance non-conditionelle $E(X_t)$ (égale à 0 ici).
La vitesse de convergence est ici inversement proportionnelle à la valeur du paramètre autorégressif ϕ.





Non stationnarité

Les tests de racine unitaire classiques (Dickey-Fuller, Phillips-Perron, KPSS, ...) montrent que l'hypothèse nulle de non stationnarité de la série est acceptée la plupart du temps.
Par conséquent, afin de stationnariser la série, l'étude est menée sur les taux de croissance ou les log-rendements de la série. Ainsi, si on observe une série $(X_t)_{t=1,\dots,T}$, la série des taux de croissance est donnée pour tout t par $Y_t = (X_t - X_{t-1})/X_{t-1}$ et la série des log-rendements est donnée pour tout t par
$R_t = \log(X_t) - \log(X_{t-1})$.

Non Normalité

- Leptokurticité
- Asymétrie

Non constance de la variance

On observe que la variance des séries subit une évolution au cours au cours du temps, en particulier sous l'effet de chocs exogènes tels que les crises financières. Ce fait empirique avéré remet alors en cause l'hypothèse d'homoscédasticité (variance constante), que l'on utilise classiquement lors d'une modélisation de série chronologique, en particulier dans le cas des processus de type ARMA. Il semble donc nécessaire de proposer des modèles prenant en compte cette hétéroscédasticité.

Agrégats de volatilité

Non seulement les séries financières ne présentent pas une variance constante au cours du temps, mais on s'aperçoit également que cette variance évolue également de manière caractéristique. En effet, les séries financières présentent des successions de phases de relative tranquillité et de phases de forte volatilité. On dit également que les séries présentent des agrégats de volatilité (*volatility clustering*).

Effet de levier

On observe une corrélation négative entre les variations des prix d'actifs et les variations de la volatilité. Toutefois, il existe une asymétrie sur les marchés dans la mesure où cette corrélation varie en intensité selon le sens de la variation des prix. En effet, on observe que la volatilité augmente fortement lorsque les prix baissent fortement (par exemple dans le cas d'une mauvaise nouvelle économique ou sur la santé financière des entreprises ou dans le cas plus général d'une crise financière). En revanche, lors des périodes d'expansion des prix, la volatilité affiche une plus grande stabilité.

Auto-Corrélations

Lorsqu'on calcule les autocorrélations des séries financières, on observe une très faible autocorrélation. Généralement, la série est blanchie par un processus AR(p)où p est relativement petit ($p \leq 3$). Il arrive même souvent que la série soit supposée suivre un bruit blanc faible (non indépendant). En revanche, les autocorrélations de la série au carré $(Y_t^2)_t$, ou élevée à une certaine puissance $(|Y_t|^\delta)_t$, présentent une forte persistance.

Co-mouvements de volatilité

Si on s'intéresse aux indices synthétiques relatifs à des marchés différents (CAC40, FTSE100, DAX, SP500, ...), on observe des mouvements de volatilité communs aux places financières, du fait d'une forte dépendance entre les marchés. En fait, les mouvements de forte volatilité s'explique par des facteurs exogènes qui s'appliquent à l'ensemble des places financières. On parle alors de co-mouvement de volatilité.

Exemples

Exemple RATS des comouvements de volatilité sur les indices boursiers de la zone euro