Session 7
Facteurs et composantes principales

MATH 60231 / 60231A
Méthodes économétriques en finance / Financial Econometrics

Anthony Sanford

HEC Montréal

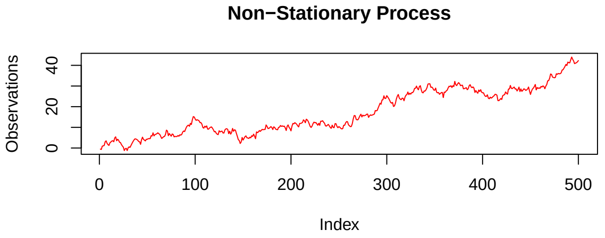
Autumn / Fall 2025

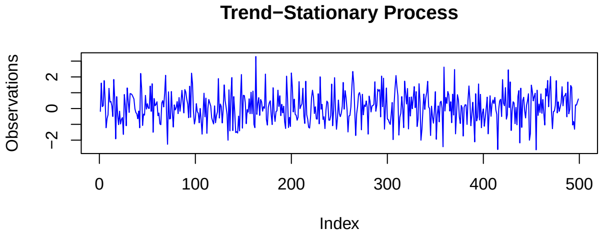
Révision rapide des modèles stationnaires, VAR et VECM I

Avant d'aller plus loin, je voulais passer quelques éléments en revue. Nous allons commencer par quelques graphiques. Les deux ensembles de graphiques suivants examineront :

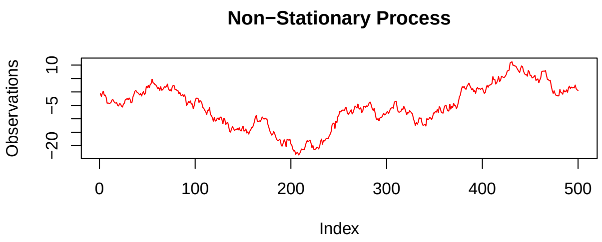
1. Un processus non stationnaire avec une tendance temporelle (processus à tendance stationnaire)
2. Un processus non stationnaire avec une tendance stochastique (un processus I(1), pour être exact)

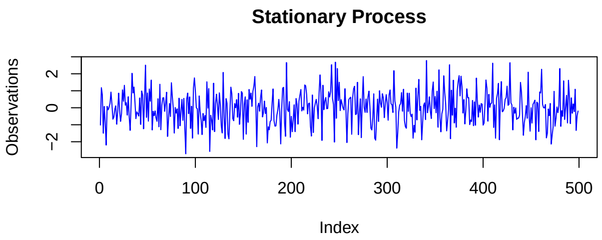
Révision rapide des modèles stationnaires, VAR et VECM II





Révision rapide des modèles stationnaires, VAR et VECM III



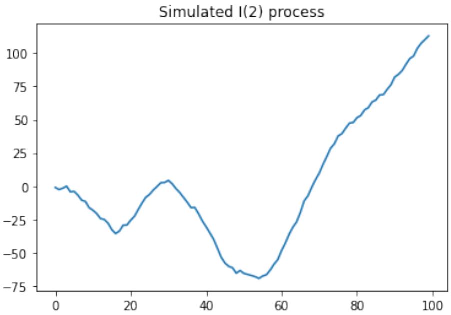


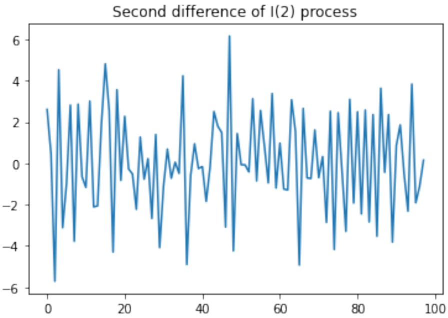
Révision rapide des modèles stationnaires, VAR et VECM IV

Un processus I(2) est une série chronologique qui doit être intégrée deux fois pour devenir stationnaire.

- Soit Y_t un processus I(2).
- On prend la première différence pour obtenir $D(Y_t) = Y_t - Y_{t-1}$, ce qui équivaut à intégrer la série une fois.
- Puisque la série originale est intégrée deux fois, nous devons prendre la première différence à nouveau pour obtenir une série stationnaire. On prend la deuxième différence pour obtenir $D^2(Y_t) = D(Y_t) - D(Y_{t-1}) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$
- La série résultante $D^2(Y_t)$ est alors stationnaire. Nous allons regarder les graphiques ensuite...

Révision rapide des modèles stationnaires, VAR et VECM V





Exemple de VAR utilisant le marché boursier canadien I

Voici un exemple de modèle VAR appliqué à un indice TSX hypothétique et au taux de change CAD/USD.

- Supposons que nous utilisons des données mensuelles de janvier 2006 à décembre 2019.
- Supposons que nous utilisons l'indice TSX comme approximation du marché canadien et le taux de change CAD/USD.
- Les données ont été générées aléatoirement, cet exemple n'utilise pas de données "réelles".
- Je ne le montre pas dans cet exemple, mais nous pourrions également utiliser le modèle VAR estimé pour ces deux variables afin de prévoir leurs valeurs futures.

Exemple de VAR utilisant le marché boursier canadien II

- Nous utiliserons un modèle VAR pour analyser la relation entre le TSX et le taux de change CAD/USD.
- Le modèle VAR peut s'écrire comme suit:

$$\mathrm{TSX}_t = c_1 + a_{11}\mathrm{TSX}_{t-1} + a_{12}\mathrm{CADUSD}_{t-1} + e_{1t}$$
$$\mathrm{CADUSD}_t = c_2 + a_{21}\mathrm{TSX}_{t-1} + a_{22}\mathrm{CADUSD}_{t-1} + e_{2t}$$

où TSX_t et CADUSD_t sont les valeurs actuelles de l'indice TSX et du taux de change CAD/USD, respectivement, et TSX_{t-1} et CADUSD_{t-1} sont leurs valeurs de la période précédente. c_1 et c_2 sont des constantes, a_{11}, a_{12}, a_{21}, et a_{22} sont des coefficients qui capturent la relation entre les variables, et e_{1t} et e_{2t} sont les termes d'erreur.

Exemple de VAR utilisant le marché boursier canadien III

Supposons que nous avons estimé le modèle à l'aide d'une régression MCO régulière comme dans la leçon 4. Le modèle estimé est alors:

$$\begin{aligned} \text{TSX}_t &= 10000 + 0.7\text{TSX}_{t-1} - 0.3\text{CADUSD}_{t-1} + e_{1t} \\ \text{CADUSD}_t &= 1.5 + 0.3\text{TSX}_{t-1} + 0.5\text{CADUSD}_{t-1} + e_{2t} \end{aligned}$$

- Ces coefficients suggèrent, par exemple, qu'une augmentation de l'indice TSX le mois précédent (TSX_{t-1}) est associée à une augmentation de l'indice TSX le mois courant (TSX_t), et qu'une diminution du taux de change CAD/USD le mois précédent (CADUSD_{t-1}) est associée à une augmentation de l'indice TSX le mois courant (TSX_t).

Exemple de VAR utilisant le marché boursier canadien IV

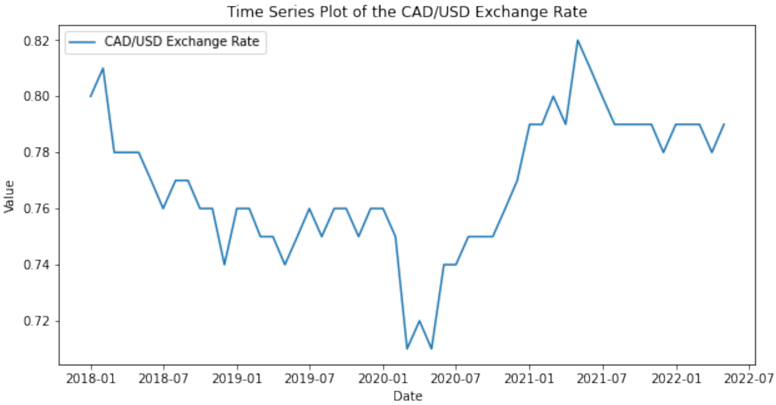


Figure: Time Series Plot of the CAD/USD Exchange Rate

Exemple de VAR utilisant le marché boursier canadien V

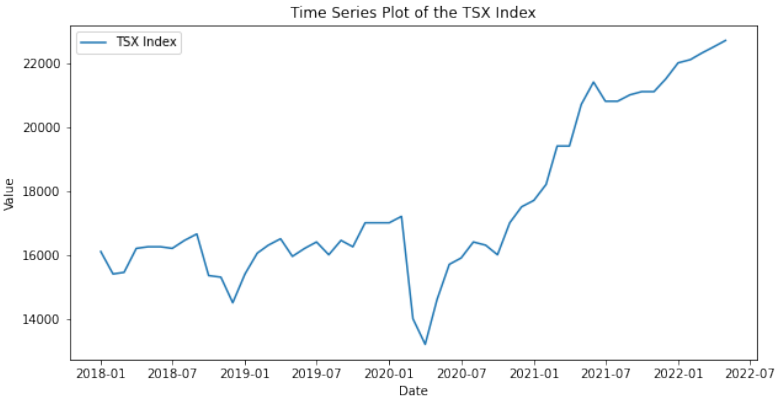


Figure: Time Series Plot of the TSX Index

Exemple de VAR utilisant le marché boursier canadien VI

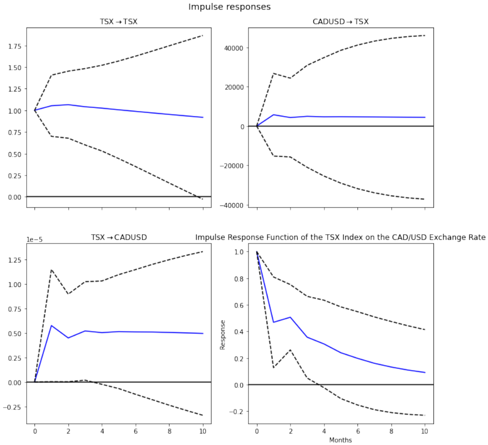


Figure: Impulse Response Functions

Exemple de VAR utilisant le marché boursier canadien VII

Il y a cependant un problème avec cette configuration. Les séries chronologiques ne semblent pas stationnaires. Ainsi, j'ai effectué des tests de racine unitaire :

- À l'aide du test ADF, nous ne parvenons pas à rejeter l'hypothèse nulle selon laquelle les deux séries chronologiques sont stationnaires.
- En utilisant le test KPSS, nous rejetons l'hypothèse nulle.

Comme nous l'avons dit dans la leçon 6, nous conclurions que la série est I(1). En tant que telle, nos coefficients seront biaisés et notre modèle est mal spécifié. Nous aurons besoin du modèle VECM pour corriger la non-stationnarité et (éventuellement) la cointégration.

Exemple de VECM

Nous aurons besoin de quelques détails de la leçon/du matériel d'aujourd'hui (modèles de composants principaux et de facteurs) pour parcourir un exemple complet du modèle VECM... donc je vais présenter un exemple de VECM dans la session 9.

Références

Nos trois manuels couvrent tous les modèles factoriels et les composantes principales.

HMPY 12.1-12.3

Tsay 9.4

Ruppert et Matteson 18.1-18.2

Nous examinerons comment appliquer cela aux modèles de la courbe des taux.

HMPY 12.5

Diebold et Li (2006)

Modèles factoriels I

Introduction

Maintenant que vous comprenez les **séries chronologiques** et comment modéliser la dynamique, nous allons nous pencher sur des problèmes d'**apprentissage non supervisé**.

- Dans notre introduction au Machine Learning, nous avons examiné l'**apprentissage supervisé** (nous avions des variables *cibles*).
- L'**apprentissage non supervisé** tente de construire des modèles simples de... tout ?
- Nous l'utilisons souvent pour réduire la *dimension* d'un problème.

Example

Taux d'intérêt : Varient selon l'échéance, l'emprunteur, les conditions

- Quels sont les plus *influents* ?
- Combien devons-nous en inclure dans un modèle ? *Lesquels* ?

Pensez à la courbe des taux... que nous utiliserons comme exemple à la fin.

Modèles factoriels II

Introduction

Les **modèles factoriels** capturent les mouvements de nombreuses séries *reliées* en les modélisant à l'aide d'un petit nombre de variables clés : les **facteurs**.

- Cela réduit la *dimensionnalité* de notre problème, le rendant plus facile à comprendre, à calculer, à prévoir, etc.
- Les **facteurs** sont ce que nous voulons modéliser / inclure dans un modèle.

Modèles factoriels III

Introduction

Je suppose que vous avez déjà vu des modèles factoriels auparavant. Par exemple, le modèle à facteurs de Fama-French explique les rendements des actions en fonction :

- du facteur de risque de marché, du facteur de taille et du facteur de valeur
- qui capturent les variations systématiques des rendements liées aux mouvements globaux du marché, à la taille d'une entreprise et à la valeur de ses actions par rapport à sa valeur comptable.

Modèles factoriels IV

Introduction

L'évaluation des actifs repose sur des modèles factoriels. La leçon 8 détaillera leur utilisation dans l'évaluation des actifs. Dans la leçon d'aujourd'hui, nous allons :

1. décrire les modèles factoriels en général
2. découvrir les composantes principales, le modèle factoriel statistique le plus utilisé.
3. examiner un exemple qui vise à utiliser les composantes principales pour modéliser la courbe des taux.

Modèles factoriels I

Les bases

Definition

Si nous avons un modèle de données de panel où les K variables explicatives sont les mêmes pour toutes les unités (donc $X_{i,t} = X_t \, \forall i$), alors c'est un **modèle factoriel**.

Example

$$y_{it} = \alpha_i + \beta_{1i} \cdot f_{1t} + \beta_{2i} \cdot f_{2t} + \ldots + \beta_{Ki} \cdot f_{Kt} + \varepsilon_{it} \tag{1}$$

où

- les f_{it} sont les K facteurs (scalaires)
- les β_{ki} sont les *charges factorielles* (scalaires)
- $\text{Cov}\left(f_{kt}, \varepsilon_{it}\right) = 0 \; \forall k, i, t$ (il s'agit donc d'une régression valide)

Modèles factoriels II

Les bases

Dans (1), les facteurs contrôlent la variation temporelle et les bêtas contrôlent la variation transversale.

Fact

Un même modèle factoriel peut être écrit de différentes manières.

Le modèle suivant est identique à (1) :

$$Y_t = \alpha + B \cdot \overrightarrow{f}_t + \varepsilon_t \tag{2}$$

où

- $Y_t \equiv [y_{1t}, \ldots, y_{Nt}]', \quad \varepsilon_t \equiv [\varepsilon_{1t}, \ldots, \varepsilon_{Nt}]', \quad \alpha \equiv [\alpha_1, \ldots, \alpha_N]'$
- $B \equiv N \times K$ matrice avec éléments β_{ik}
- $\overrightarrow{f}_t \equiv [f_{1t}, \ldots, f_{Kt}]'$

Modèles factoriels III

Les bases

Le modèle suivant est identique à (1) *et* (2) :

$$Y_i = 1_T \cdot \alpha_i + F \cdot \overrightarrow{\beta}_i + \varepsilon_i \tag{3}$$

où

- 1_T est un vecteur $T \times 1$ composé de 1
- F est une matrice $T \times K$ de facteurs
- ε_i est un vecteur $T \times 1$ de termes d'erreur
- $\overrightarrow{\beta}_i$ est un vecteur $K \times 1$ de charges factorielles

Fact

Le MEDAF (ou CAPM) est un modèle à un facteur pour les prix des actifs.

$$(r_t^i - r_t^f) = \beta_i \cdot (r_t^m - r_t^f) + \varepsilon_t$$

Modèles à facteurs latents

Lorsqu'on considère les modèles factoriels comme des régressions de données de panel, on envisage de régresser des Y sur des X (les facteurs) pour estimer les β (les charges factorielles).

Mais les modèles factoriels sont également utilisés *lorsque nous n'observons pas les* X. C'est ce qu'on appelle un **modèle à facteurs latents**.

Il existe plusieurs façons d'estimer les modèles à facteurs latents. Nous allons examiner la plus répandue : les **composantes principales**.

Composantes Principales I

Les bases

Supposons que nous avons une collection de N variables $Y_t \equiv [y_{1t} \ldots y_{Nt}]'$ que nous voulons "résumer."

Problem

Nous souhaitons choisir un ensemble de N pondérations $\alpha \mid \alpha' \cdot \alpha = 1$ tout en maximisant $\text{Var}(\alpha' Y_t)$

- Notez que $\text{Var}(\alpha' Y_t) = \alpha' \cdot \Sigma \cdot \alpha$ où $\Sigma \equiv \text{Var}(Y_t)$

Solution

$\alpha \equiv v_1$, *le premier vecteur propre de* Σ.

Definition

$PC_{1t} \equiv v_1' Y_t$ est la **première composante principale** de Y_t

Composantes Principales II

Les bases

Problem

Que faire si PC_{1t} ne "suffit" pas ? Comment obtenir des facteurs supplémentaires ?

Pourquoi pas :

1. Régresser (MCO) Y_t sur PC_{1t} pour obtenir les résidus ϵ_t ?
2. Créer une nouvelle CP à partir de ϵ_t ?

Remark: Ceci est équivalent à

Problem

$$\max_{\alpha_2} \mathrm{Var}\left(\alpha_2' Y_t\right) \quad s.t. \quad \alpha_2' \cdot \alpha_2 = 1 \text{ et } \alpha_2' \cdot \alpha_1 = 0$$

Solution

$\alpha_2 = v_2$, *le second vecteur propre de* Σ.

$PC_{2t} \equiv \alpha_2' Y_t \equiv v_2' Y_t$ *est la* **deuxième composante principale** *de* Y_t

Composantes Principales III

Les bases

Nous pouvons répéter cette astuce jusqu'à ce que nous ayons "assez" de facteurs.

Problem

$$\max_{\alpha_j} \mathrm{Var}\left(\alpha_j' Y_t\right) \quad s.t. \quad \alpha_j' \cdot \alpha_j = 1 \text{ et } \alpha_j' \cdot \alpha_i = 0 \ \forall i < j$$

Solution

$\alpha_j = v_j$, *le j-ième vecteur propre de* Σ.
$PC_{jt} \equiv \alpha_j' Y_t \equiv v_j' Y_t$ *est la* **j-ième composante principale** *de* Y_t

Nous pouvons créer jusqu'à N CP de cette manière.

Composantes Principales IV

Les bases

Les composantes principales ont plusieurs propriétés particulières.[1]

1. $\text{Var}(PC_{it}) = \lambda_i$, la i-ième valeur propre de Σ.
   Ainsi $\text{Var}\left(PC_{1t}\right) \geq \text{Var}\left(PC_{2t}\right) \geq \ldots \geq \text{Var}\left(PC_{Nt}\right)$
   - ... car les résidus ne cessent de diminuer !
2. $\text{Cov}(PC_{it}, PC_{jt}) = 0$ sauf si $i = j$
   Les CP ne sont pas corrélées entre elles !
   - ... car nous travaillons toujours avec des résidus MCO !
3. Non seulement les CP sont des combinaisons linéaires des N différents Y, mais les Y peuvent être écrits comme des combinaisons linéaires des N CP !

La dernière condition signifie que l'on peut **décomposer** les mouvements de toute série Y_{it} en mouvements des N CP **et vice-versa !**

[1] Ces propriétés découlent des propriétés des valeurs propres et vecteurs propres des matrices symétriques. Voir l'annexe de ces diapositives ou le chapitre A.20 de Ruppert et Matteson pour en savoir plus.

Composantes Principales

Décompositions et pondérations

Pour comprendre ce que chaque PC_j capture,
nous pouvons examiner les pondérations attribuées à chaque série Y_i.

- Les pondérations sont données par le vecteur propre v_j.

Pour comprendre les mouvements de chaque série Y_i,
nous pouvons examiner les pondérations attribuées à chaque PC_j.

- La pondération de Y_i sur PC_j est $\beta_{ij} \equiv [v_j]_i$
  où $[v_j]_i$ est le i-ième élément du j-ième vecteur propre.

Nous pouvons également utiliser cette équivalence pour décomposer la variance de nos Y_t.

- $\therefore Var(Y_t) = \sum_{i=1}^{N} \lambda_i \cdot v_i v_i'$
- $\therefore \sum_{i=1}^{N} Var(Y_{it}) = \sum_{i=1}^{N} \lambda_i$

HMPY donne un exemple détaillé à ce sujet aux pages 357 à 359.

Diagrammes d'éboulis ("Scree plots") I

Étant donné que :

1. Y_{it} peut être décomposé en une somme pondérée des N PC_{jt}, et
2. Chacun des PC_{jt} a une variance de λ_j,

Nous pouvons :

1. Résumer la variance totale de $\mathbf{Y}_t$ par $\sum_{j=1}^{N} \lambda_j$, et
2. Calculer la **fraction de la variance** expliquée par les j premiers CP comme : $\left(\sum_{i=1}^{j} \lambda_i\right) / \left(\sum_{i=1}^{N} \lambda_i\right)$.

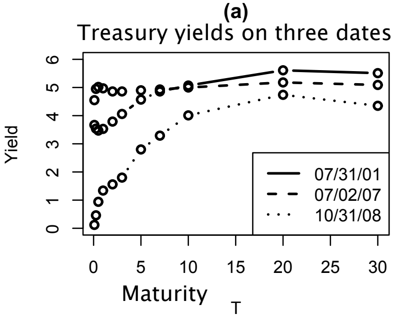
Diagrammes d'éboulis ("Scree plots") Graphiques représentant j par rapport à λ_j ou à $\lambda_j / \left(\sum_{i=1}^{N} \lambda_i\right)$

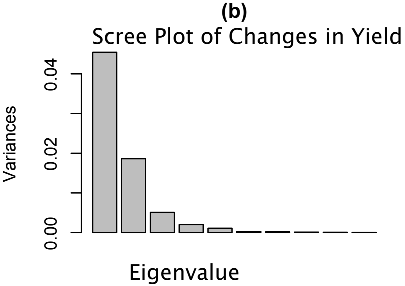
Example

Ruppert et Matteson calculent des CP pour les rendements quotidiens d'obligations du Trésor américain avec 11 échéances différentes (1990-2008).

Diagrammes d'éboulis ("Scree plots") II

```
Importance of components:
                          PC1  PC2   PC3   PC4   PC5    PC6
Standard deviation      0.21 0.14 0.071 0.045 0.033 0.0173
Proportion of Variance 0.62 0.25 0.070 0.028 0.015 0.0041
Cumulative Proportion  0.62 0.88 0.946 0.974 0.989 0.9932
```





Un exemple numérique simple I

Nous disposons d'un ensemble de données avec deux variables, la taille (en pouces) et le poids (en livres) de 5 individus :

Taille (po)	Poids (lbs)
63	150
67	160
71	170
68	165
72	180

Un exemple numérique simple II

Pour effectuer l'ACP, nous standardisons d'abord les données en soustrayant la moyenne et en divisant par l'écart-type de chaque variable afin de les standardiser, car elles ne sont pas sur la même échelle. Cela nous donne :

Taille (po)	Poids (lbs)
-1,46	-1,30
-0,24	-0,65
0,98	0,01
0,12	-0,32
0,61	2,26

Un exemple numérique simple III

Nous calculons ensuite la matrice de covariance des données standardisées :

$$\begin{pmatrix} 1,00 & 0,851 \\ 0,851 & 1,00 \end{pmatrix}$$

Les éléments diagonaux représentent les variances de chaque variable, tandis que les éléments hors diagonale représentent la covariance entre les deux variables.

Un exemple numérique simple IV

Les vecteurs propres et les valeurs propres de la matrice de covariance sont :

$$\begin{pmatrix} 0,71 & -0,71 \\ 0,71 & 0,71 \end{pmatrix} \begin{pmatrix} 1,85 & 0 \\ 0 & 0,15 \end{pmatrix} \begin{pmatrix} 0,71 & 0,71 \\ -0,71 & 0,71 \end{pmatrix}$$

Vous pouvez les obtenir à l'aide d'un calculateur de vecteurs propres et de valeurs propres.

Un exemple numérique simple V

Les valeurs de la CP1 et de la CP2 pour chaque individu sont :

Individu	CP1	CP2
1	-1,76	0,91
2	-0,42	0,21
3	0,49	-0,01
4	-0,10	-0,24
5	1,00	-0,97

Comment avons-nous obtenu ces valeurs ?

Un exemple numérique simple VI

- Rappelons que le premier vecteur propre (v1) est : $v1 = (0,71, 0,71)$.
- Pour projeter les données standardisées sur v1, nous prenons le produit scalaire de chaque point de données (sous forme standardisée) avec v1. Par exemple, pour le premier point de données (-1,46, -1,30), nous avons : $CP1 = (-1,46 * 0,71) + (-1,30 * 0,71) = -1,76$.
- Nous faisons cela pour chaque point de données et obtenons une valeur de CP1 pour chaque individu.
- En prenant la moyenne de ces valeurs, nous obtenons la valeur de CP1 pour l'ensemble des données : $CP1 = (-1,76 + (-0,42) + 0,49 + (-0,10) + 1,00)/5 = 0,02$
- Ainsi, la première composante principale (CP1) pour cet ensemble de données est 0,02.

Un exemple numérique simple VII

Nous faisons la même chose pour la CP2:

- Rappelons que le deuxième vecteur propre (v2) est : $v2 = (-0,71, 0,71)$
- Pour projeter les données standardisées sur v2, nous prenons le produit scalaire de chaque point de données (sous forme standardisée) avec v2. Par exemple, pour le premier point de données (-1,46, -1,30), nous avons : $CP2 = (-1,46 * (-0,71)) + (-1,30 * 0,71) = 0,91$.
- Nous faisons cela pour chaque point de données et obtenons une valeur de CP2 pour chaque individu.
- En prenant la moyenne de ces valeurs, nous obtenons la valeur de CP2 pour l'ensemble des données : $CP2 = (0,91 + 0,21 - 0,01 - 0,24 - 0,97)/5 = -0,11$
- Ainsi, la deuxième composante principale (CP2) pour cet ensemble de données est -0,11.

Un exemple numérique simple VIII

Nous pouvons également voir la proportion de la variance expliquée par chacune des CP. Nous pouvons la calculer en utilisant les valeurs propres : 1,85 (pour CP1) et 0,15 (pour CP2).

- Pour calculer la proportion de la variance expliquée par chaque composante principale, nous divisons chaque valeur propre par la somme totale des valeurs propres : $\lambda_1 + \lambda_2 = 2,00$
- Proportion CP1 $=\lambda_1$ / somme des lambda = 1,85/2 = 0,925
- Proportion CP2 = λ_2 / somme des lambda = 0,15/2 = 0,075

La première composante principale (CP1) explique 92,5% de la variance totale des données, et la deuxième composante principale (CP2) explique 7,5% de la variance totale.

Un exemple numérique simple IX

Lorsque nous effectuons une ACP sur cet ensemble de données, nous constatons que la première composante principale (CP1) représente la direction dans laquelle les données varient le plus.

- Dans ce cas, CP1 représente la taille corporelle globale, ce qui n'est pas surprenant puisque la taille et le poids sont tous deux liés à la taille corporelle.
- Le fait que CP1 explique la plus grande partie de la variation des données suggère que la taille corporelle est le facteur le plus important pour déterminer les différences entre ces individus.

Un exemple numérique simple X

La deuxième composante principale (CP2) représente la direction dans laquelle les données varient le moins.

- Dans ce cas, CP2 pourrait représenter la proportion du poids qui est due à la masse musculaire par rapport à la masse graisseuse.
- Le fait que CP2 explique moins de variations dans les données que CP1 suggère que les différences dans la teneur en muscle par rapport à la graisse ne sont pas aussi importantes que la taille corporelle globale pour déterminer les différences entre ces individus.
- Comment savons-nous que CP2 pourrait représenter la masse musculaire par rapport à la masse graisseuse ?

Un exemple numérique simple XI

La réponse est que nous ne le savons pas. Nous déduisons en fonction de ce que nous voyons dans les résultats.

- Nous pouvons voir, cependant, que v2 a une valeur négative pour la taille et une valeur positive pour le poids. Par conséquent, nous pouvons dire que v2 représente une combinaison de la taille et du poids qui est liée à la proportion du poids qui est due à la masse musculaire par rapport à la masse graisseuse.
- Le fait que v2 ait des valeurs à la fois positives et négatives pour la taille et le poids indique que les personnes qui sont plus grandes et plus lourdes ou plus petites et plus légères peuvent avoir des valeurs similaires de CP2, en fonction de la proportion de leur poids qui est due aux muscles par rapport à la graisse.
- Mais encore une fois, à ce stade, cela n'est peut-être pas clair, et nous pourrions avoir besoin de construire des facteurs...

Des composantes principales à un modèle factoriel I

Pour construire un modèle statistique, on peut simplement utiliser les $K < N$ premières composantes principales comme facteurs

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta_{1i} \cdot PC_{1t} + \beta_{2i} \cdot PC_{2t} + \ldots + \beta_{Ni} \cdot PC_{Nt} \\
&= \alpha_i + \beta_{1i} \cdot PC_{1t} + \beta_{2i} \cdot PC_{2t} + \ldots + \beta_{Ki} \cdot PC_{Kt} + \varepsilon_{it} \\
&= \alpha_i + \beta_{1i} \cdot f_{1t} + \beta_{2i} \cdot f_{2t} + \ldots + \beta_{Ki} \cdot f_{Kt} + \varepsilon_{it} \qquad (4)
\end{aligned}
$$

où $\varepsilon_{it} \equiv \sum_{j=K+1}^{N} \beta_{ji} \cdot PC_{jt}$ reflète les CP que nous avons supprimées du modèle.

(En évaluation d'actifs, nous pourrions considérer ϵ_{it} comme un "risque idiosyncratique." Voir HMPY p. 355.)

Des composantes principales à un modèle factoriel II

omment choisir K ?

Ad hoc Examiner les diagrammes d'éboulis et décider quelle quantité est "suffisante."

Bai et Ng Choisir K pour *minimiser* un critère d'information.

$$IC_K = log(\hat{V}_K) + K \cdot \left(\frac{N+T}{N \cdot T}\right) \cdot log\left(\frac{N \cdot T}{N+T}\right)$$

$$\hat{V}_K = \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2$$

où $\hat{u}_{it}$ sont les variances résiduelles d'un modèle à K facteurs.

Des composantes principales à un modèle factoriel III

Bien que nous ayons défini les composantes principales en fonction de la matrice de covariance Σ, nous construisons souvent des modèles factoriels en utilisant la matrice de **corrélation**.

- Cela équivaut à diviser d'abord chaque série par son écart type, puis à calculer la matrice de covariance.

Cela peut être important si nos variables sont mesurées dans des unités différentes.

- Les CP sont choisies pour expliquer le plus de variance possible.
- Si nous mesurons une variable en cm, elle aura une variance plus grande que si nous la mesurons en km. (Par un facteur de 100000^2).
- Cela peut fausser les CP et les amener à expliquer ou non cette variable.

Example

En appliquant l'ACP pour étudier les fluctuations économiques américaines, Stock et Watson (2003) utilisent la matrice de corrélation.

Retour en arrière : Le test de Johansen pour la cointégration I

Rappel sur la cointégration :

- La cointégration se produit lorsque deux ou plusieurs séries chronologiques non stationnaires ont une relation à long terme. Cela signifie qu'elles évoluent ensemble dans le temps, malgré des fluctuations potentiellement importantes à court terme.
- Cela peut représenter un défi pour les modèles VAR car les modèles VAR standards supposent que toutes les variables du système sont stationnaires.
- Si la cointégration est présente dans un modèle VAR, les paramètres estimés pourraient être biaisés ou incohérents, conduisant à des inférences incorrectes sur les relations sous-jacentes entre les variables.

Retour en arrière : Le test de Johansen pour la cointégration II

Le test le plus puissant pour la cointégration est le test de **Johansen**. Il permet de tester :

- La présence ou l'absence de cointégration.
- Le nombre de relations distinctes de cointégration.

Remark: L'intuition derrière le test de Johansen est basée sur la relation entre la cointégration et les composantes principales.

- Dans le VECM, le rang de la matrice Π indique le nombre de tendances stochastiques *distinctes*.
- Le rang d'une matrice carrée est donné par le nombre de valeurs propres $\lambda > 0$.
- Cela doit donc aussi nous indiquer le nombre de CP statistiquement significatives !

Retour en arrière : Le test de Johansen pour la cointégration III

Le test de Johansen pour la cointégration est en fait composé de deux tests différents. Les deux utilisent les valeurs propres ordonnées $(\lambda_1 > \lambda_2 > \ldots > \lambda_N)$ de $\hat{\Pi}$.

1. Le test de **Trace** vérifie s'il y a *plus* que r_0 vecteurs de cointégration indépendants ; $H_0 : r = r_0 \quad vs \quad H_A : r > r_0$ pour un r_0 spécifié par l'utilisateur, $0 \leq r_0 < N$

$$LR_{Trace} \equiv -T \cdot \sum_{r_0+1}^{N} \ln(1 - \lambda_i) \tag{5}$$

2. Le test de la **valeur propre maximale** vérifie s'il existe des preuves de *un* vecteur de cointégration indépendant supplémentaire ; $H_0 : r = r_0 \quad vs \quad H_A : r = r_0 + 1$

$$LR_{Max} \equiv -T \ln(1 - \lambda_{r_0+1}) \tag{6}$$

Ces tests vérifient le nombre de CP significatives **en niveaux**.

Aperçu d'un exemple

Contexte : Supposons que nous avons trois variables dans un VECM :

- Y_t : PIB
- π_t : Inflation
- i_t : Taux d'intérêt

L'objectif est de déterminer le nombre de relations de cointégration (r) entre ces variables.

Hypothèses du test de Johansen :

- Hypothèse nulle (H_0) : $r \leq r_0$ (nombre de vecteurs de cointégration)
- Hypothèse alternative (H_1) : $r > r_0$

Résultats du test de Johansen

Valeurs propres et statistiques de test :

Hypothèse	Test Trace	Valeur critique	Test Max-Eigen	Valeur critique
$r = 0$	50.3	29.8	31.2	21.1
$r = 1$	19.1	15.4	12.8	14.0
$r = 2$	6.3	3.8	6.3	3.8

Table: Résultats du test de Johansen

Interprétation des résultats

Test Trace (cumulatif) :

- Pour $r = 0$: La statistique Trace ($50.3 > 29.8$) rejette l'hypothèse nulle. Il existe au moins un vecteur de cointégration.
- Pour $r = 1$: $19.1 > 15.4$, donc nous rejetons l'hypothèse nulle d'au plus un vecteur. Il y en a au moins deux.
- Pour $r = 2$: $6.3 > 3.8$, donc nous rejetons l'hypothèse nulle d'au plus deux vecteurs. Cela implique $r = 3$, c'est-à-dire trois relations de cointégration.

Test Max-Eigen (marginal) :

- Pour $r = 0$: $31.2 > 21.1$, donc nous rejetons l'hypothèse nulle. Ajouter un vecteur de cointégration est significatif.
- Pour $r = 1$: $12.8 < 14.0$, donc nous ne rejetons pas l'hypothèse nulle. Pas d'évidence d'un deuxième vecteur de cointégration.

Concilier les tests

Principaux constats :

- Le test Trace suggère $r = 3$: toutes les trois variables sont liées par trois relations de cointégration.
- Le test Max-Eigen suggère $r = 1$: il y a au moins une relation de cointégration.

Implications pratiques :

- Si $r = 3$, le VECM doit inclure trois termes de correction d'erreur, un pour chaque relation de cointégration.
- Chaque terme de correction d'erreur explique les écarts par rapport à un équilibre de long terme.
- Ce modèle capture les dynamiques d'ajustement entre les trois variables pour maintenir l'équilibre.

Relations de long terme dans le VECM

Termes de correction d'erreur :

- Puisque $r = 3$, il existe trois relations de cointégration :

$$\beta_{11} Y_t + \beta_{12}\pi_t + \beta_{13} i_t, \quad \beta_{21} Y_t + \beta_{22}\pi_t + \beta_{23} i_t, \quad \beta_{31} Y_t + \beta_{32}\pi_t + \beta_{33} i_t$$

- Ces relations représentent des combinaisons linéaires stationnaires des variables.
- Les écarts par rapport à ces équilibres entraînent des ajustements dans le VECM.

Dynamiques dans le VECM :

$$\Delta Y_t = \gamma_{11}(\mathrm{ECT}_1) + \gamma_{12}(\mathrm{ECT}_2) + \gamma_{13}(\mathrm{ECT}_3) + \text{dynamiques court terme}$$

ECT_i : Terme de correction d'erreur correspondant à la relation i.
$\gamma_{11}, \gamma_{12}, \gamma_{13}$: Vitesses d'ajustement vers l'équilibre.

Diebold and Li (2006) I

Prévision de la structure par terme des taux d'intérêt

- La structure par terme des taux d'intérêt décrit la variation des taux d'intérêt dans le temps pour différents types d'obligations ayant des échéances différentes.
- Diebold et Li (2006) ont proposé une méthode de prévision de la structure par terme, basée sur un modèle à facteurs latents.
- Le modèle décompose la structure par terme en quelques facteurs sous-jacents.

Diebold and Li (2006) II

Prévision de la structure par terme des taux d'intérêt

Le modèle suppose que la structure par terme peut être représentée comme une combinaison linéaire de facteurs, plus un terme de bruit :

$$y_t = \alpha_t + \beta_t' f_t + \epsilon_t$$

- y_t est un vecteur des rendements obligataires au temps t.
- α_t est un vecteur d'interceptions.
- β_t est un vecteur de pondérations de facteurs.
- f_t est un vecteur de facteurs latents.
- ϵ_t est un vecteur d'erreurs idiosyncratiques.

Les facteurs latents sont estimés à l'aide de **l'analyse en composantes principales (ACP)**, qui identifie les combinaisons linéaires qui expliquent le plus de variations dans le modèle.

Diebold and Li (2006) III

Prévision de la structure par terme des taux d'intérêt

- Ces facteurs peuvent ensuite être utilisés pour prévoir la structure future des taux d'intérêt, en extrapolant leurs valeurs actuelles et en ajoutant les variations prévues des interceptions et des chargements factoriels.
- Le modèle fournit un cadre de travail utile pour comprendre la structure par terme des taux d'intérêt et prévoir ses mouvements futurs.

Cela devrait vous donner une intuition pour le modèle... nous allons examiner des détails plus précis en utilisant une notation plus spécifique

Diebold and Li (2006) IV

Prévision de la structure par terme des taux d'intérêt

Diebold et Li (2006) proposent un modèle factoriel simple pour la courbe des taux :

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \cdot \left(\frac{1 - e^{-\lambda \cdot \tau}}{\lambda \cdot \tau} \right) + \beta_{3t} \cdot \left(\frac{1 - e^{-\lambda \cdot \tau}}{\lambda \cdot \tau} - e^{-\lambda \cdot \tau} \right) + \varepsilon_t(\tau) \qquad (7)$$

où

- $y_t(\tau) \equiv$ le rendement à l'échéance au temps t d'une obligation ayant une échéance τ.
- $\beta_{1t}, \beta_{2t}, \beta_{3t} \equiv$ des **facteurs** variant dans le temps qui capturent les mouvements de la courbe des taux.
- $\lambda \equiv 0.0609$.

Diebold and Li (2006) V

Prévision de la structure par terme des taux d'intérêt

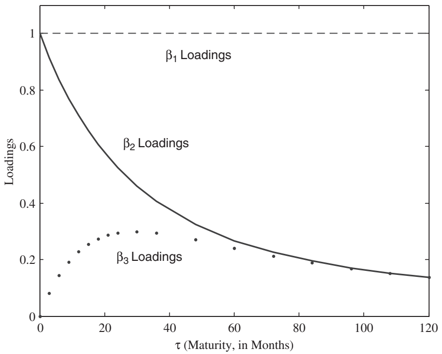


Fig. 1. Factor loadings. We plot the factor loadings in the three-factor model,

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau}\right),$$

where the three factors are β_{1t}, β_{2t}, and β_{3t}, the associated loadings are 1, $(1 - e^{-\lambda_t \tau})/\lambda_t \tau$, and $(1 - e^{-\lambda_t \tau})/\lambda_t \tau - e^{-\lambda_t \tau}$, and τ denotes maturity. We fix $\lambda_t = 0.0609$.

Session 7 Facteurs et composantes principales

Anthony Sanford

Diebold and Li (2006) VI

Prévision de la structure par terme des taux d'intérêt

1. β_{1t} est un facteur de "niveau" ou de "long terme" :
   Son poids $= 1 \iff$ influence toutes les échéances de manière égale $\iff$ déplace le niveau de la courbe des taux
   (y compris les échéances les plus longues).
2. β_{2t} est un facteur de "pente" ou de "court terme" :
   Son poids $\left(\frac{1-e^{-\lambda\cdot\tau}}{\lambda\cdot\tau}\right) \to 1$ lorsque $\tau \to 0$ et $\to 0$ lorsque $\tau \to \infty$.
   Il capture principalement les mouvements des rendements à court terme.
3. β_{3t} est un facteur de "courbure" ou de "moyen terme" :
   Son poids $\left(\frac{1-e^{-\lambda\cdot\tau}}{\lambda\cdot\tau} - e^{-\lambda\cdot\tau}\right) \to 0$ lorsque $\tau \to 0$ et lorsque $\tau \to \infty$, mais est strictement positif entre les deux.

Des valeurs différentes de $(\beta_{1t}, \beta_{2t}, \beta_{3t})$ peuvent produire des courbes de rendement en pente ascendante, descendante, présentant des bosses, des bosses inversées, etc.

Diebold and Li (2006) VII

Prévision de la structure par terme des taux d'intérêt

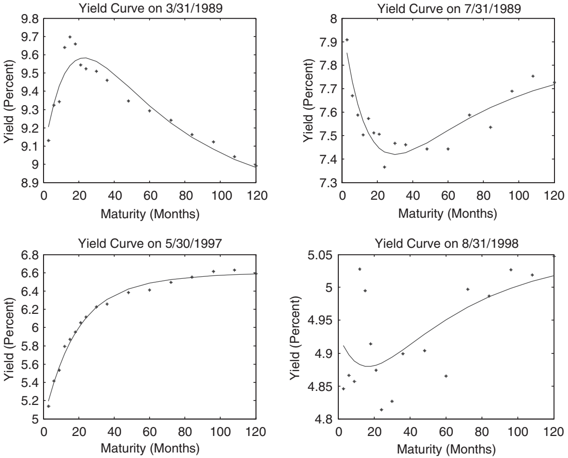


Fig. 5. Selected fitted (model-based) yield curves. We plot fitted yield curves for selected dates, together with actual yields. See text for details.

Diebold and Li (2006) VIII

Prévision de la structure par terme des taux d'intérêt

Alors, comment pouvons-nous estimer (7) ?

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \cdot \left(\frac{1 - e^{-\lambda \cdot \tau}}{\lambda \cdot \tau} \right) + \beta_{3t} \cdot \left(\frac{1 - e^{-\lambda \cdot \tau}}{\lambda \cdot \tau} - e^{-\lambda \cdot \tau} \right) + \varepsilon_t(\tau)$$

1. $\lambda \equiv 0,0609$, donc les coefficients de chaque facteur sont connus pour chaque valeur de τ.
2. Diebold et Li (2006) effectuent T régressions MCO, chacune estimant $\{\beta_{1t}, \beta_{2t}, \beta_{3t}\}$ en utilisant des données sur $N = 17$ échéances.
3. Ils notent que $\{\widehat{\beta}_{1t}, \widehat{\beta}_{2t}, \widehat{\beta}_{3t}\}$ ont de faibles corrélations entre eux, mais des autocorrélations élevées.

Diebold and Li (2006) IX

Prévision de la structure par terme des taux d'intérêt

Diebold et Li comparent les estimations de leurs 3 facteurs à deux alternatives.

1. Les trois premières composantes principales.
2. Des mesures simples du niveau, de la pente et de la courbure.
   - Niveau = rendement à 10 ans
   - Pente = rendement à 10 ans - rendement à 3 mois
   - Courbure = 2 x rendement à 2 ans - (rendement à 3 mois + rendement à 10 ans)

"...our estimated factors are close to the first three principal components." [p. 306]

Diebold and Li (2006) X

Prévision de la structure par terme des taux d'intérêt

Fig. 7. Model-based level, slope and curvature (i.e., estimated factors) vs. data-based level, slope and curvature. We define the level as the 10-year yield, the slope as the difference between the 10-year and 3-month yields, and the curvature as the twice the 2-year yield minus the sum of the 3-month and 10-year yields.

Lectures complémentaires I

(Facultatif)

L'application de l'ACP nécessite d'estimer la matrice de covariance (ou de corrélation) Σ de taille $N \times N$.

- **$N \ll T$** Facile.
- **$N < T$** Plus difficile. $\widehat{\Sigma}$ peut manquer de précision, conduisant à des estimations imprécises des pondérations et des facteurs.
- **T important, N plus important** Nos estimations de Σ n'auront pas un rang complet ! Cette matrice possède $\frac{1}{2} \cdot N \cdot (N+1)$ éléments et nous n'avons que $N \cdot T$ observations.

Conner et Korajczyk (1986) proposent une méthode alternative pour les grands N appelée "Composantes Principales Asymptotiques". Voir Tsay Section 9.6 pour une description.

Lectures complémentaires II

(Facultatif)

L'application de la décomposition en valeurs propres à la matrice de covariance n'est pas la seule façon d'utiliser les composantes principales pour construire un modèle factoriel statistique.

- Toute combinaison linéaire de nos composantes principales équivalente à une rotation à N dimensions fonctionnera également. Voir Tsay 9.5.2 ou Ruppert et Matteson 18.6.

Il n'y a rien qui *exige* que les facteurs de Nelson-Siegel ressemblent aux composantes principales. La similitude entre eux s'est rompue après 2007, lorsque de nombreux taux d'intérêt sont tombés à près de zéro.