

M2 Advanced Econometrics

Samuel Ligonnière

Année universitaire 2024-2025

M2 Master Innovation, Marchés et Science des Données -
Université Evry Paris-Saclay

Introduction générale

Cours basé sur une étude de cas général. Etude typiquement sollicitée par une institution publique nationale or internationale.

Comment prédire les variations économiques et financières?

1. Présentation des cycles économiques et financiers
2. Explication des différentes variables économiques, dans les régimes économiques, régimes monétaires, et régimes financiers.
3. Comment prédire les futures fluctuations?

Introduction générale: le type de données

(1) Données en séries temporelles, $N=1$, T grand

- Problème de la fréquence d'observations: toutes les observations en période T disponibles? Ou semaine pour les uns, mois pour les autres?
- Distribution des données: prix d'une cotation boursière en continu. Quelle valeur choisir?

Séries temporelles utiles pour résoudre les problèmes suivants:

- Comment la valeur de l'indice boursier d'un pays a varié en fonction de ses fondamentaux macroéconomiques.
- Comment le cours de l'action d'une entreprise a évolué lorsqu'elle a annoncé la valeur de son dividende.
- L'effet d'une augmentation du déficit commercial d'un pays sur son taux de change.

Introduction générale: le type de données

(2) Données en coupe transversale, N grand, T=1

- Un sondage sur l'utilisation des services de courtage en ligne.
- Une coupe transversale des rendements des actions au NYSE.
- Un échantillon de notations de crédit obligataire pour les banques britanniques.

Cross-section utilisable pour répondre notamment aux questions suivantes:

- La relation entre la taille d'une entreprise et le rendement de l'investissement dans ses actions.
- La relation entre le niveau de PIB d'un pays et la probabilité que le gouvernement fasse défaut sur sa dette souveraine.

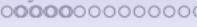
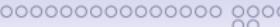
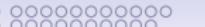
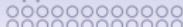
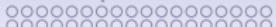
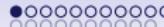
Introduction générale: le type de données

(3) Données de panel, N grand, T grand

- Permettent de traiter un plus large éventail de problèmes et d'aborder des questions complexes.
- Facilitent l'analyse dynamique des variables et de leurs relations, en augmentant les degrés de liberté et la puissance des tests, tout en réduisant les problèmes de multicolinéarité.
- Réduisent le biais dû aux variables omises en structurant correctement le modèle.
- Attention : cette approche suppose que les valeurs moyennes des variables et leurs relations restent constantes dans le temps et entre les unités de coupe transversale.

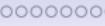
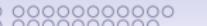
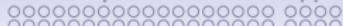
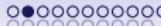
La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. ROC - Application à Dexia
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire



La boîte à outils économétriques

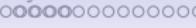
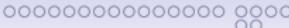
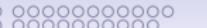
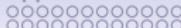
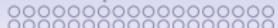
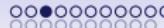
- Théorème Gauss-Markov
- Tests et p-value
- Applications
- Choix du modèle: biais, variance des estimateurs et multicolinéarité
- Variables qualitatives - Application aux crises bancaires
- ROC - Application à Dexia
- Filtres temporels
- Vers le NLP - Application à la politique monétaire



Théorème Gauss-Markov

Il existe 5 hypothèses, **H1 à H5**, qui sont collectivement appelées les hypothèses de Gauss-Markov. Elles garantissent que les estimateurs des MCO sont les meilleurs estimateurs linéaires sans biais (BLUE).

Nous utiliserons le symbole X pour représenter l'ensemble des variables indépendantes (X_1, X_2, \dots, X_k). Dans l'équation de Mincer sur le salaire avec educ, exper et CDI comme variables explicatives, nous avons $X = (\text{educ}, \text{exper}, \text{CDI})$.



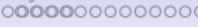
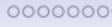
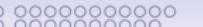
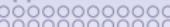
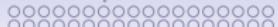
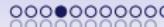
Rappel des hypothèses

H1: Le modèle est linéaire en paramètres. Dans le modèle, nous disposons de Y , X , et e tel que

$$Y = b_0 + b_1 X + e \quad (1)$$

→ La relation peut s'écrire de cette formulation suivante. Etant donné les possibilités de non-linéarité, cette formulation n'est pas si excessive qu'à première vue.

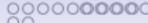
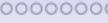
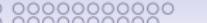
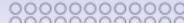
H2: Echantillon aléatoire. Nous disposons d'un échantillon iid (indépendant et identiquement distribué) de taille N , généré par le modèle de l'hypothèse H1.



Rappel des hypothèses

H3: Les X_i varient dans l'échantillon.

H4: $\mathbb{E}[e|X] = 0 = \mathbb{E}[e]$ Pour tout i , $\mathbb{E}[e_i|X_i] = 0$ On retrouve les hypothèses de base dans le chapitre sur l'espérance conditionnelle des résidus, sans biais selon une partition de la population.

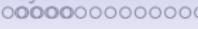
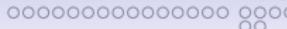
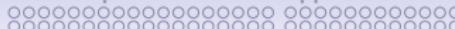


Rappel des hypothèses

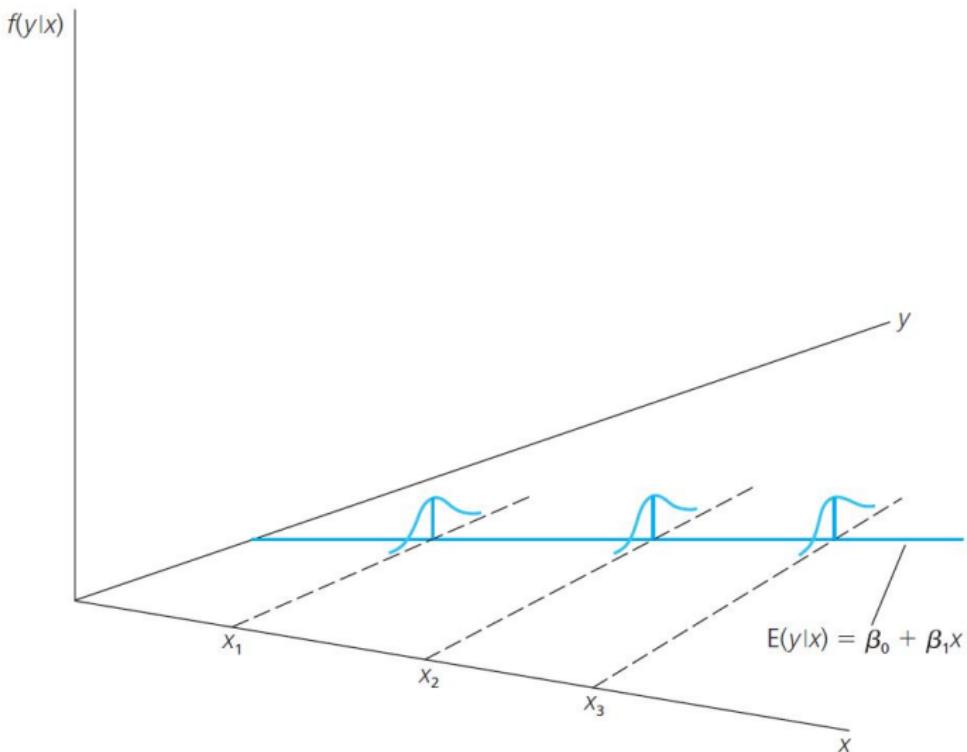
H5: Homoscédasticité. Le terme d'erreur e a la même variance, quel que soit la valeur de X . En d'autres mots, nous avons

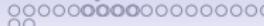
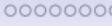
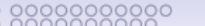
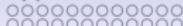
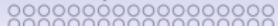
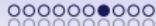
$$\text{Var}(e) = \sigma^2 \quad (2)$$

Remarque: H5 ne joue aucun rôle sur le caractère sans biais des estimateurs.



Homoscédasticité des résidus





Rappel des hypothèses

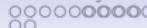
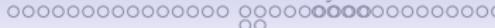
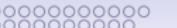
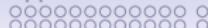
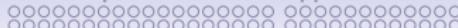
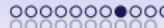
Nous pouvons alors écrire les hypothèses **H1** et **H4** comme suit :

$$E(Y | X) = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_K X_K \quad (3)$$

L'hypothèse **H5** stipule que :

$$\text{Var}(Y | X) = \sigma^2 \quad (4)$$

Autrement dit, la variance de Y , conditionnée par X , ne dépend pas des valeurs des variables indépendantes.



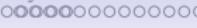
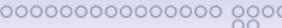
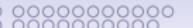
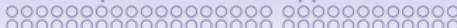
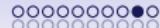
Modèle de régression multiple (forme matricielle)

Pour démontrer Gauss-Markov, nous allons d'abord exprimer le modèle de régression multiple sous forme matricielle, ce qui a plein d'avantages calculatoires :

$$Y = X\beta + e \quad (5)$$

où :

- Y est un vecteur $N \times 1$ des observations de la variable dépendante,
- X est une matrice $N \times (k + 1)$ des observations des k variables explicatives (y compris l'ordonnée à l'origine),
- β est un vecteur $(k + 1) \times 1$ des coefficients inconnus (paramètres à estimer),
- e est un vecteur $N \times 1$ des termes d'erreur.

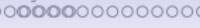
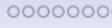
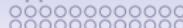
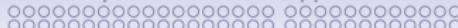
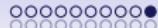


Exemple concret

Supposons que nous ayons $N = 5$ observations et deux variables explicatives X_1 (éducation) et X_2 (expérience). Voici les matrices associées :

Le vecteur Y est :

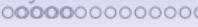
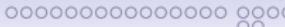
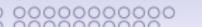
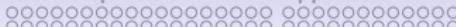
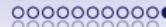
$$Y = \begin{pmatrix} 35 \\ 42 \\ 50 \\ 48 \\ 55 \end{pmatrix}$$



Matrice X (variables explicatives)

La matrice X , avec une colonne de 1 pour l'ordonnée à l'origine, et les deux variables explicatives X_1 (éducation) et X_2 (expérience), est :

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ 1 & X_{31} & X_{32} \\ 1 & X_{41} & X_{42} \\ 1 & X_{51} & X_{52} \end{pmatrix}$$



Vecteur β (coefficients)

Le vecteur β des coefficients inconnus est :

$$\beta = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

Modèle matriciel complet

Le modèle complet sous forme matricielle est donc :

$$\begin{pmatrix} 35 \\ 42 \\ 50 \\ 48 \\ 55 \end{pmatrix} = \begin{pmatrix} 1 & 12 & 5 \\ 1 & 15 & 6 \\ 1 & 16 & 7 \\ 1 & 14 & 8 \\ 1 & 18 & 9 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

Ce qui revient à

$$\underbrace{Y}_{N \times 1} = \underbrace{X}_{N \times (k+1)} \underbrace{\beta}_{(k+1) \times 1} + \underbrace{e}_{N \times 1}$$

Estimateur des Moindres Carrés Ordinaires (MCO)

L'estimateur des MCO pour β est donné par :

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (6)$$

où X' est la transposée de la matrice X , et $(X'X)^{-1}$ est l'inverse de la matrice $X'X$.

→ Démonstration dans les slides suivantes

Démonstration : Calcul de $\hat{\beta}$

Nous cherchons à minimiser :

$$S(\beta) = e'e = (Y - X\beta)'(Y - X\beta)$$

Développons cette expression :

$$S(\beta) = (Y' - \beta'X')(Y - X\beta)$$

Ce qui donne :

$$S(\beta) = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

En remarquant que $Y'X\beta$ et $\beta'X'Y$ sont des scalaires et donc égaux (i.e. $Y'X\beta = \beta'X'Y$), nous pouvons réécrire :

$$S(\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

Pour minimiser cette expression par rapport à β , nous prenons la dérivée partielle de $S(\beta)$ par rapport à β , et nous l'égalisons à zéro :

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'Y + 2X'X\beta = 0$$

En simplifiant, nous obtenons l'équation normale :

$$X'X\beta = X'Y$$

Enfin, en multipliant par $(X'X)^{-1}$ (si $X'X$ est inversible), nous obtenons l'estimateur des MCO :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

Quelles propriétés des estimateurs MCO à tester?

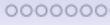
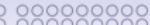
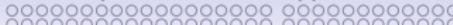
Sous les hypothèses **H1** à **H4** (linéarité, échantillonnage aléatoire, absence de multicolinéarité parfaite, et espérance conditionnelle nulle des erreurs), l'espérance de $\hat{\beta}$ est donnée par :

$$E(\hat{\beta}) = (X'X)^{-1}X'E(Y)$$

Sachant que $E(Y) = X\beta$, nous obtenons :

$$E(\hat{\beta}) = (X'X)^{-1}X'X\beta = \beta$$

Donc, $\hat{\beta}$ est un estimateur sans biais de β . → Qu'en est-il de la variance?



Variance? (1/4)

Par définition, la variance de $\hat{\beta}$ sous forme matricielle est définie comme :

$$\text{Var}(\hat{\beta}) = E \left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right] \quad (7)$$

En substituant $\hat{\beta} = (X'X)^{-1}X'Y$ dans l'équation ci-dessus, nous obtenons :

$$\hat{\beta} - \beta = (X'X)^{-1}X'(Y - X\beta) \quad (8)$$

Remplaçons Y par $X\beta + e$ (modèle linéaire) :

$$\hat{\beta} - \beta = (X'X)^{-1}X'(X\beta + e) - \beta$$

Variance? (2/4)

Décomposons et simplifions l'expression obtenue précédemment :

$$\hat{\beta} - \beta = (X'X)^{-1}X'(X\beta) + (X'X)^{-1}X'e - \beta$$

1. Simplification du premier terme :

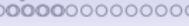
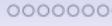
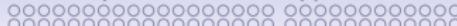
$$(X'X)^{-1}X'(X\beta) = \beta$$

2. Finalisation : En regroupant, nous obtenons :

$$\hat{\beta} - \beta = \beta + (X'X)^{-1}X'e - \beta$$

$$\hat{\beta} - \beta = (X'X)^{-1}X'e$$

L'écart entre $\hat{\beta}$ et β est donc lié uniquement aux résidus e .



Variance? (3/4)

Maintenant que nous savons que :

$$\hat{\beta} - \beta = (X'X)^{-1}X'e$$

La variance de $\hat{\beta}$ est donnée par :

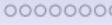
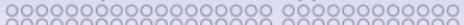
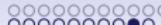
$$\text{Var}(\hat{\beta}) = \text{Var}((X'X)^{-1}X'e)$$

Propriété clé : Si A est une matrice constante, alors :

$$\text{Var}(Ae) = A\text{Var}(e)A'$$

En appliquant cette propriété :

$$\text{Var}(\hat{\beta}) = (X'X)^{-1}X'\text{Var}(e)X(X'X)^{-1}$$



Variance? (4/4)

Sous l'hypothèse d'homoscédasticité (**H5**), la variance des termes d'erreur e est constante, c'est-à-dire que :

$$\text{Var}(e) = \sigma^2 I_n \quad (9)$$

où I_n est la matrice identité de taille N et σ^2 est la variance constante des erreurs. En substituant cette expression dans notre équation de variance, nous obtenons :

$$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} X' I_n X (X'X)^{-1} \quad (10)$$

Comme $X' I_n X = X'X$, cela simplifie à :

$$\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \quad (11)$$

Récapitulatif sur la variance

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad (12)$$

Cela signifie que la variance de l'estimateur des MCO dépend de la variance des erreurs σ^2 et de la matrice $(X'X)^{-1}$, qui capture les relations entre les variables explicatives.

Non-démontré: variance de l'estimateur MCO est plus faible que la variance de tout autre estimateur. Estimateur MCO est donc BLUE.

Gauss-Markov
oooooooooooo

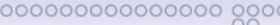
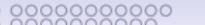
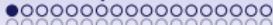
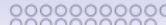
Tests et p-value
oooooooooooo

Applications
oooooooooooo

Choix du modèle
oooooooooooo

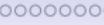
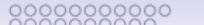
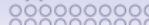
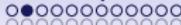
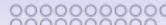
Dummies
oooooooooooo

ROC Cycles
oooooooooooo



La boîte à outils économétriques

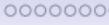
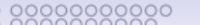
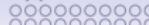
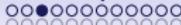
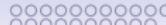
1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. ROC - Application à Dexia
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire



Distribution d'échantillonnage et hypothèse de normalité

Pour simplifier, on suppose que l'erreur non observée est distribuée normalement. Il s'agit de notre **H6**: l'erreur e est indépendante des variables explicatives x_1, x_2, \dots, x_k et suit une loi normale.

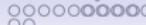
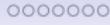
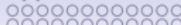
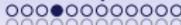
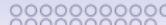
$$e \sim \text{Normal}(0, \sigma^2) \quad (13)$$



Hypothèses du Modèle Linéaire Classique (MLC)

H1 à H6 forment les hypothèses du modèle linéaire classique (MLC). Les estimateurs des MCO sont les estimateurs non biaisés de variance minimale parmi tous les estimateurs (pas seulement linéaires).

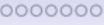
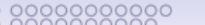
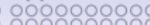
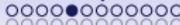
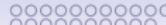
$$y|x \sim \text{Normal}(b_0 + b_1x_1 + \cdots + b_kx_k, \sigma^2) \quad (14)$$



Problèmes de normalité dans certaines applications

Distribution normale de e parfois peu plausible

- Le salaire, conditionnellement au niveau d'éducation, d'expérience, n'a pas de raison d'être normalement distribué. Pas de salaire négatif non plus. Le salaire minimum, s'il existe, va aussi à l'encontre de cette distribution normale.
- Dans certains cas, la transformation logarithmique peut aider, comme $\log(\text{prix à la consommation})$
- Problème de non-normalité du terme d'erreur e mineur face aux grands échantillon

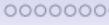
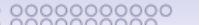
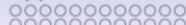
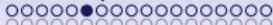
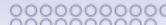


Distribution d'échantillonnage des estimateurs

Sous les hypothèses **H1** à **H6** :

$$\hat{b}_j \sim \text{Normal}(b_j, \text{Var}(\hat{b}_j))$$

$$(\hat{b}_j - b_j)/\text{sd}(\hat{b}_j) \sim \text{Normal}(0, 1)$$



Conclusion

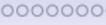
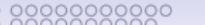
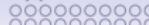
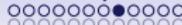
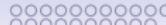
- La normalité de l'erreur entraîne la normalité des distributions d'échantillonnage des estimateurs MCO.
- La non-normalité dans de grands échantillons n'est pas problématique.

Distribution t des estimateurs standardisés

Pour construire les tests, avec les hypothèses du MLC (**H1 à H6**), nous avons

$$(\hat{b}_j - b_j) / \text{se}(\hat{b}_j) \sim t_{N-K-1}$$

- $\text{se}(\hat{b}_j)$ remplace σ dans la variance estimée. standard error (erreur-type) par rapport aux estimateurs.
- $K + 1$ est le nombre de paramètres inconnus, N le nombre d'observations. $N - K - 1$ nombre de degrés de liberté.
- Ce résultat est essentiel pour tester les hypothèses sur b_j .
- Rappel: différence entre écart-type (standard deviation) et erreur-type (standard error)
 - Ecart-type: mesure la dispersion des données autour de la moyenne d'un ensemble de données.
 - Erreur-type: estimation de l'écart-type de la distribution d'un estimateur.

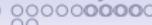
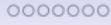
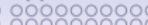
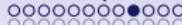
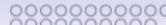


Calcul de la statistique t

Statistique t pour tester $H_0 : b_j = 0$:

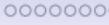
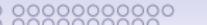
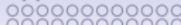
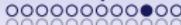
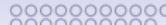
$$t_{\hat{b}_j} = \frac{\hat{b}_j}{\text{se}(\hat{b}_j)}$$

Si $|t_{\hat{b}_j}|$ est grand, on rejette H_0 . Rejet de H_0 si $|t_{\hat{b}_j}| > t_{\alpha, N-K-1}$



Règle de décision et niveau de significativité

- On fixe un niveau de significativité (α) pour décider si $|t_{\hat{b}_j}|$ est suffisamment éloigné de zéro. α est généralement à 10, 5 ou encore 1%.
- Distribution de $t_{\hat{b}_j}$ sous H_0 : t_{N-K-1}
- Rejet de H_0 en fonction de la valeur réciproque de $t_{\hat{b}_j}$ et de $t_{\alpha, N-K-1}$



Test unilatéral à droite

Nous allons faire ce test $H_0 : b_j = 0$ contre l'hypothèse alternative $H_1 : b_j > 0$. Nous comparons ainsi la statistique t et la valeur critique.

$$t_{\hat{b}_j} = \frac{\hat{b}_j}{\text{se}(\hat{b}_j)} >< t_{\alpha, N-K-1}$$

Règle de décision: rejet de H_0 si $|t_{\hat{b}_j}| > t_{\alpha, N-K-1}$

- $t_{\alpha, N-K-1}$ valeur critique, déterminant le niveau de risque de type I (probabilité de rejeter H_0 quand H_0 vrai).
- α est le niveau de significativité du test.

Test unilatéral à droite et rejet de H₀

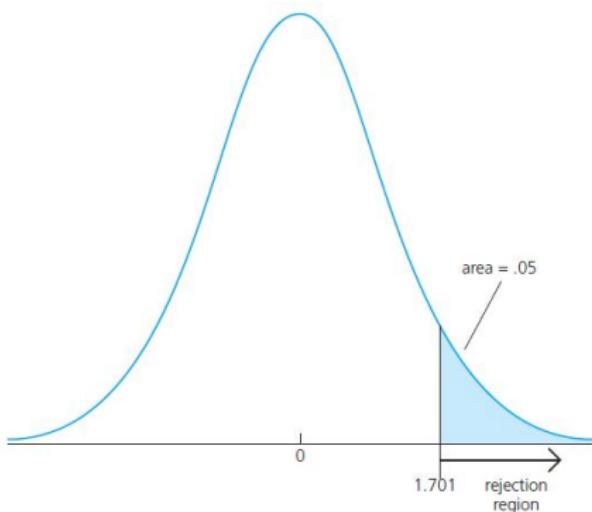


Table de Student pour trouver cette valeur critique, de 1,701

- Illustration de cette zone de rejet, α à 5%
- Degrés de liberté: $N - K - 1 = 28$

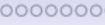
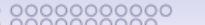
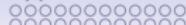
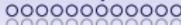
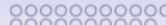
Table de Student $1 - \alpha = 0.95$; $N - K - 1 = 28$

La distribution t de Student

$$\Pr[T \leq t] = \int_{-\infty}^t \frac{\Gamma((r+1)/2)}{\sqrt{\pi r} \Gamma(r/2)} \frac{1}{(1+x^2/r)^{(r+1)/2}} dx$$

$$\Pr[T \leq -t] = 1 - \Pr[T \leq t]$$

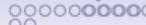
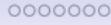
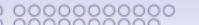
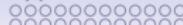
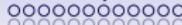
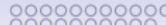
r	$\Pr[T \leq t]$				
	0.90	0.95	0.975	0.99	0.995
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
75	1.293	1.665	1.992	2.377	2.643
100	1.290	1.660	1.984	2.364	2.626
200	1.286	1.653	1.972	2.345	2.601
500	1.283	1.648	1.965	2.334	2.586
1000	1.282	1.646	1.962	2.330	2.581
∞	1.282	1.645	1.960	2.326	2.576



Valeurs critiques

2 réflexions:

- Plus le niveau de significativité exigé est élevé (i.e. α faible), plus la valeur critique sera forte.
- Plus les degrés de liberté sont importants (i.e. très grand nombre d'observations N), plus la valeur critique est basse.
Cela se rapproche des valeurs critiques de la loi normale (1,645 pour $\alpha = 5\%$, et 2,326 pour $\alpha = 1\%$)

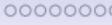
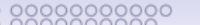
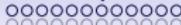
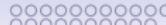


Valeurs critiques

2 types d'erreurs à dissocier:

1. Erreur de type I: H_0 est vrai, alors que le test rejette H_0
2. Erreur de type II: H_0 est faux, alors que le test ne rejette pas H_0

Critère plus strict de rejection (i.e. α faible)? Rejet plus rare de l'hypothèse nulle, réduisant les erreurs de type I mais augmentant les erreurs de type II.

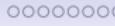
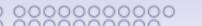
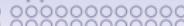
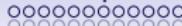
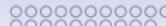


Test bilatéral

Test $H_0 : b_j = 0$ contre l'hypothèse alternative $H_1 : b_j \neq 0$.

Ce test est souvent plus prudent, étant donné que nous pouvons ignorer le sens d'une relation entre 2 variables, a priori.

Néanmoins, la zone de rejet concerne les deux extrémités de la distribution. Si nous considérons le seuil à 5%, cela signifie qu'il faut considérer 2,5% à gauche et 2,5% à droite.



Test bilatéral

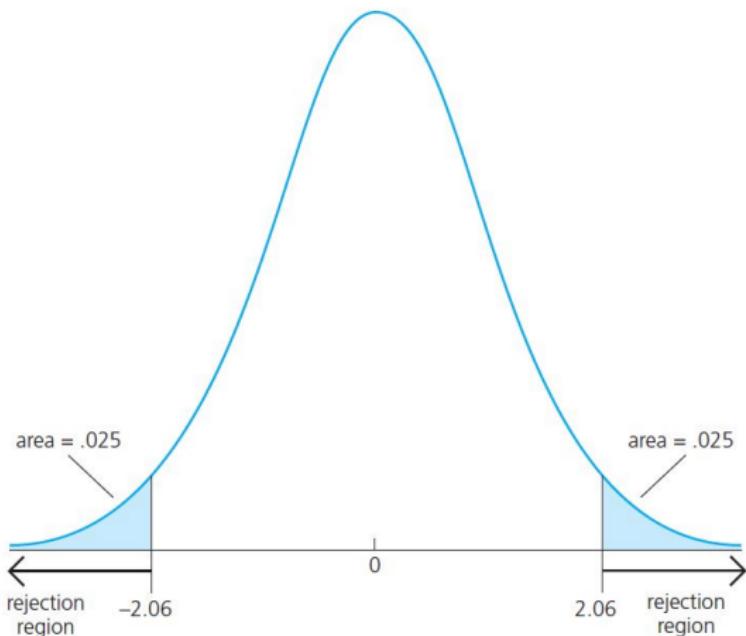
La règle de décision pour rejet de H_0 , au niveau de α , est alors

$$|t_{\hat{b}_j}| > t_{1-\frac{\alpha}{2}, N-K-1} \quad (15)$$

ce qui est équivalent à

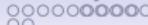
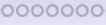
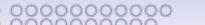
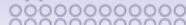
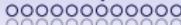
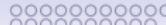
$$t_{\hat{b}_j} > t_{1-\frac{\alpha}{2}, N-K-1} \quad \text{ou si} \quad t_{\hat{b}_j} < t_{\frac{\alpha}{2}, N-K-1} = -t_{1-\frac{\alpha}{2}, N-K-1} \quad (16)$$

Test bilatéral



Cas pour $N - K - 1 = 25$, et pour chaque côté, on cherche
 $1 - \alpha = 0,975$.

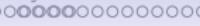
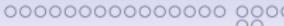
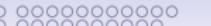
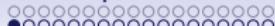
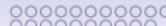
Samuel Ligonièvre



Pourquoi la p-value?

Jusqu'à présent, la méthodologie était

- Choix d'un niveau de significativité (α) et détermination d'une valeur critique.
- Comparaison de la statistique t avec la valeur critique : rejet ou non de H_0 à un niveau de significativité donné.
- Limite : la sélection d'un niveau de significativité peut être arbitraire.



Pourquoi la p-value?

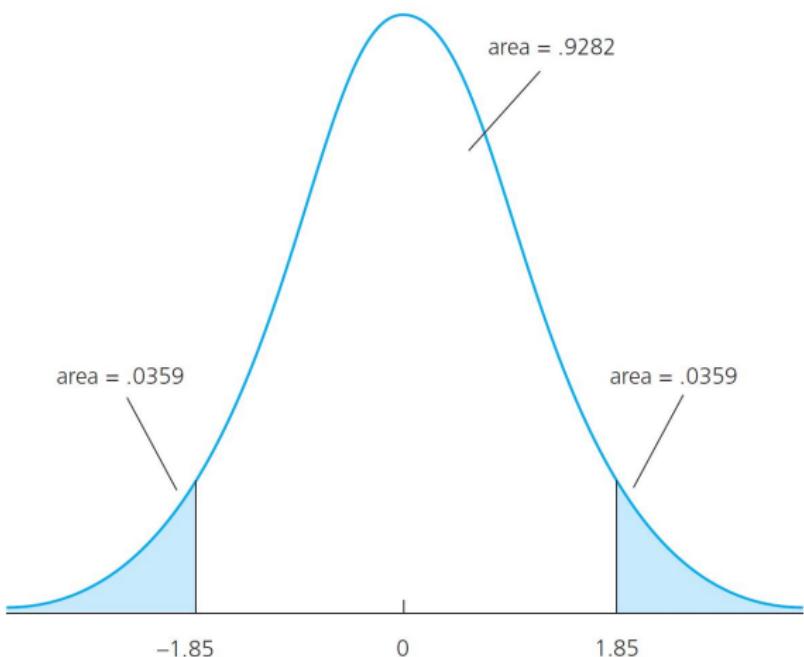
Prenons le cas d'une t-stat à 1,85 et 40 degrés de liberté, avec un test bilatéral,

- Au niveau α de 5% (donc divisé par 2), la valeur critique est à 2,021, donc on ne rejette pas H_0 .
- Au niveau α de 10% (donc divisé par 2), la valeur critique est à 1,684, donc on rejette H_0 .

Définition de la p-value

La p-value est le plus petit niveau de significativité auquel H_0 serait rejeté pour une valeur donnée de la statistique t .

Illustration de la p-value



Cas pour $N - K - 1 = 40$, avec $t = 1,85$.

Samuel Ligoni  re

(Ann  e universitaire 2024-2025)

M2 Advanced Econometrics

48 / 185

Calcul et interprétation de la p-value

Dans cet exemple avec 40 degrés de liberté et $t = 1.85$:

$$\text{p-value} = 2 \cdot P(T > 1.85) = 2 \cdot 0.0359 = 0.0718$$

Interprétation : si H_0 est vrai, on observerait une statistique t aussi extrême que 1.85 environ 7.2% du temps.

Conclusion : une p-value de 0.0718 fournit une certaine évidence contre H_0 , mais pas assez pour rejeter H_0 au niveau de 5%.

Utilisation de la p-value dans les tests classiques

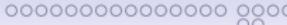
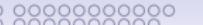
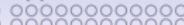
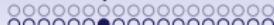
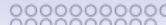
Une fois la p-value calculée, un test classique peut être réalisé à tout niveau de signification souhaité. Règle : on rejette H_0 si $p\text{-value} < \alpha$.

La p-value représente la probabilité d'observer une statistique t aussi extrême que celle obtenue, si H_0 est vrai. Plus la p-value est faible, plus on rejette avec force H_0 .

P-value sur test unilatéral possible. Pour obtenir la p-value unilatérale à partir de la p-value bilatérale, il suffit de diviser par 2.

Règles générales

1. L'importance de la valeur de p-value: dépend de la taille de l'échantillon. La p-value peut parfois être acceptable à 0,20 avec un très faible N, p-value à 0,001 avec un très grand N.
2. Possibilité que des variables avec des faibles t-stat soient avec les *mauvais* signes pour les coefficients. Mais tout cela, ce sont généralement des variables non-statistiquement significatifs. En d'autres termes, "-0.001 non-significatif signifie zéro".
3. La significativité statistique est importante, mais l'importance économique l'est tout autant: exercice de quantification.



Test sur une combinaison de paramètres

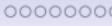
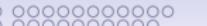
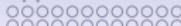
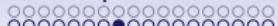
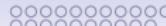
Par exemple, test avec hypothèse nulle : $H_0 : \beta_1 = \beta_2$, soit

$$H_0 : \beta_1 - \beta_2 = 0$$

Face à l'hypothèse alternative :

$$H_1 : \beta_1 < \beta_2$$

Utilisation d'une statistique t pour tester H_0 en vérifiant si la différence estimée $\hat{\beta}_1 - \hat{\beta}_2$ est statistiquement significative.



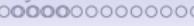
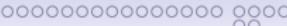
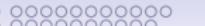
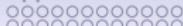
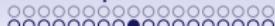
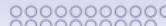
Calcul de la statistique t

T-stat pour tester $H_0 : \beta_1 - \beta_2 = 0$:

$$t = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\text{erreur-type}(\widehat{\beta}_1 - \widehat{\beta}_2)}$$

La difficulté réside dans le calcul de erreur-type($\widehat{\beta}_1 - \widehat{\beta}_2$).

$$\text{Var}(\widehat{\beta}_1 - \widehat{\beta}_2) = \text{Var}(\widehat{\beta}_1) + \text{Var}(\widehat{\beta}_2) - 2\text{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) \quad (17)$$



Introduction au Test F

La statistique t permet de tester si un coefficient MCO est égal à une constante donnée (souvent zéro).

Cependant, pour tester des hypothèses impliquant plusieurs restrictions, on utilise le test F.

Exemple : déterminer si un ensemble de variables a un effet significatif sur la variable dépendante.

Exemple de restrictions d'exclusion

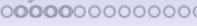
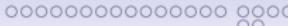
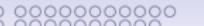
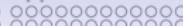
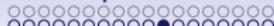
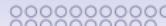
Modèle pour expliquer le salaire des joueurs de football

$$\log(\text{Salary}) = \beta_0 + \beta_1 \cdot \text{Years} + \beta_2 \cdot \text{Intern} + \beta_3 \cdot \text{NbrMatchs} + \beta_4 \cdot \text{Buts} + e$$

Hypothèse nulle : le nombre d'années de jeu et le fait d'être un international ne jouent pas sur le salaire, après avoir contrôlé pour les 4 déterminants.

$$H_0 : \beta_1 = 0, \quad \beta_2 = 0$$

Hypothèse alternative : au moins l'un des coefficients β_1 ou β_2 est différent de zéro. Cela correspond à 2 restrictions d'exclusion ici. Cela s'appelle un test d'hypothèses multiples, également test d'hypothèses jointes.



Exemple de restrictions d'exclusion

Modèle non restreint (inclus toutes les variables) :

$$\log(\text{Salary}) = \beta_0 + \beta_1 \cdot \text{Years} + \beta_2 \cdot \text{Intern} + \beta_3 \cdot \text{NbrMatchs} + \beta_4 \cdot \text{Buts} + e$$

Modèle restreint (exclut les variables d'années et le statut international) :

$$\log(\text{Salary}) = \beta_0 + \beta_3 \cdot \text{NbrMatchs} + \beta_4 \cdot \text{Buts} + e$$

Calcul du Test F avec les SSR

On utilise la somme des résidus au carré (SSR) pour comparer les deux modèles.

Modèle non restreint :

$$\log(\text{salary}) = 11.19 + 0.0689 \cdot \text{Years} + 0.0126 \cdot \text{Intern} \\ + 0.00098 \cdot \text{NbrMatchs} + 0.0144 \cdot \text{Buts}$$

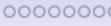
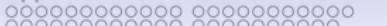
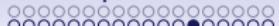
(0.29) (0.0121) (0.0026) (0.00110) (0.0161)

$N = 353, \quad \text{SSR} = 183,186, \quad R^2 = 0,6278$

Modèle restreint :

$$\log(\text{salary}) = 11.22 + 0.0713 \cdot \text{NbrMatchs} + 0.0202 \cdot \text{Buts} \\ (0.11) (0.0125) (0.0013)$$

$N = 353, \quad \text{SSR} = 198,311, \quad R^2 = 0,5971$



Formule de la F-stat

Le test F utilise l'augmentation de la SSR pour déterminer si les restrictions sont valides :

$$F = \frac{(SSR_{restreint} - SSR_{non\ restreint}) / q}{SSR_{non\ restreint} / (N - k - 1)}$$

Où :

- SSR est la somme des carrés des résidus du modèle,
- q est le nombre de restrictions (également la différence de degrés de liberté entre les 2 modèles),
- N est la taille de l'échantillon,
- k est le nombre de variables explicatives dans le modèle non restreint.

La F-stat est toujours non-négative, car

$$SSR_{restreint} \geq SSR_{non\ restreint}$$

Samuel Ligoni  re

Formule de la F-stat

Le test F permet de tester des restrictions multiples de manière efficace en comparant les modèles restreint et non restreint. Le test est basé sur l'augmentation de la SSR du modèle restreint, et il indique si les variables exclues ont un effet significatif sur la variable dépendante.

La F-stat est censée suivre la distribution suivante: $F \sim F_{q, N-k-1}$

L'hypothèse nulle H_0 est d'autant plus facilement rejetée que F suffisamment "grand", dépendant du niveau de significativité retenue.

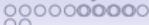
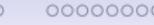
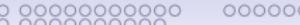
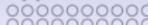
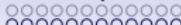


Table avec $\alpha = 5\%$

		Numerator Degrees of Freedom									
		1	2	3	4	5	6	7	8	9	10
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
D	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
e	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
n	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
m	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
a	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
t	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
o	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
r	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
D	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
e	21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
g	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
r	23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
e	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
s	25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
o	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
f	27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
F	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
r	29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
e	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
d	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
o	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
m	90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
m	120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
m	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Dans notre cas, avec $\alpha = 5\%$, la valeur critique nous est donnée par $F_{2,353-4-1} = 3$

Samuel Ligonièvre

Table avec $\alpha = 1\%$

Numerator Degrees of Freedom											
	1	2	3	4	5	6	7	8	9	10	
D	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
e	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
n	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
o	13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
m	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
j	15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
n	16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
a	17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
t	18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
r	19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
D	20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
e	21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
g	22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
r	23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
e	24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
s	25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
o	26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
f	27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
F	28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
r	29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
e	30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
d	40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
o	60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
m	90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
m	120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
	∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Dans notre cas, avec $\alpha = 1\%$, la valeur critique nous est donnée par $F_{2,353-4-1} = 4.61$

Samuel Ligoni  re

Relation entre F et T-statistiques

- Une restriction conjointe incluant $b_1 = 0$ aura moins de pouvoir pour détecter $b_1 \neq 0$ que la statistique t habituelle.
- Exemple : regrouper une variable significative avec des variables non significatives pourrait masquer la signification statistique.

Souvent, lorsqu'une variable est très significative statistiquement, le groupe de variables testé conjointement sera également significatif. Dans ces cas, il n'y a pas de contradiction logique à rejeter les deux hypothèses nulles.

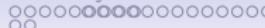
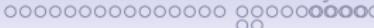
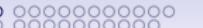
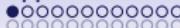
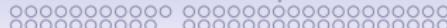
Forme R^2 de la F-stat

Pour tester les restrictions d'exclusion, il est souvent plus pratique d'utiliser la forme de R^2 de la statistique F, basée sur les R^2 des modèles restreint (r) et non restreint (nr).

En utilisant $SSR_r = SST(1 - R_r^2)$ et $SSR_{nr} = SST(1 - R_{nr}^2)$, la statistique F devient :

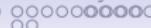
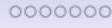
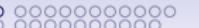
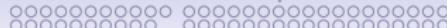
$$F = \frac{(R_{nr}^2 - R_r^2)/q}{(1 - R_{nr}^2)/(N - k - 1)}$$

Cela simplifie le calcul puisque R^2 est toujours entre 0 et 1, contrairement aux SSR qui peuvent être influencés par l'unité de mesure de Y .



La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. **Applications**
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. ROC - Application à Dexia
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire



Applications

De multiples applications en économie et en finance existent:

- CAPM
- Peut-on battre le marché? Jensen (1968)
- L'hypothèse de sur-réaction du marché DeBondt & Thaler (1985, 1987)

Exemple (1): CAPM - MEDAF

Frontière Moyenne-Variance (Markowitz, 1959)

- Les investisseurs choisissent leur portefeuille en fonction du rendement espéré (récompense) et de la variance (risque) des rendements.
- Un portefeuille efficient moyenne-variance maximise le rendement espéré pour un niveau donné de variance.
- Objectif : trouver un portefeuille avec le meilleur compromis entre risque et rendement.

Modèle d'Évaluation des Actifs Financiers (CAPM)

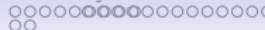
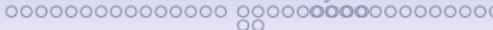
Hypothèses :

- Les investisseurs ont des attentes homogènes.
- Ils détiennent des portefeuilles efficients moyenne-variance.
- Absence de frictions de marché.

Résultat principal :

$$E[Y_{i,t}] = r_f + \beta_{im} (E[Y_{m,t}] - r_f)$$

- r_f : taux sans risque.
- $Y_{m,t}$: rendement du portefeuille de marché.
- $\beta_{im} = \frac{\text{Cov}(Y_{i,t}, Y_{m,t})}{\text{Var}(Y_{m,t})}$: sensibilité au marché.



Rendements Excédentaires

- Rendement excédentaire : $Z_{i,t} = Y_{i,t} - r_f$.
- Formulation compacte :

$$E[Z_{i,t}] = \beta_{im} E[Z_{m,t}]$$

- $Z_{m,t}$: rendement excédentaire du portefeuille de marché.
- Le CAPM établit une règle linéaire pour la tarification des actifs.

Tests empiriques du CAPM

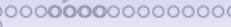
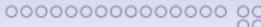
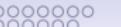
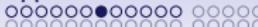
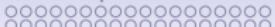
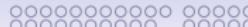
Trois implications principales :

1. La constante est nul ($\alpha = 0$).
2. La variation des rendements excédentaires entre les actifs est entièrement capturée par les bêta.
3. La prime de risque du marché $E[Z_{m,t}]$ est positive.

Régression empirique :

$$Z_{i,t} = \beta_{im} Z_{m,t} + \varepsilon_{i,t}$$

- Tests basés sur des estimateurs MCO (Moindres Carrés Ordinaires) pour le coefficient de pente (β).
- Les études empiriques utilisant des indices boursiers comme proxy du marché rejettent souvent les implications du CAPM.



Résumé sur CAPM - MEDAF

- La frontière moyenne-variance offre une base théorique pour la gestion des portefeuilles.
- Le CAPM propose une relation linéaire entre le risque systématique (β) et le rendement attendu.
- Les tests empiriques montrent des limites dans les hypothèses et les prédictions du CAPM.

Applications

De multiples applications en économie et en finance existent:

- CAPM
- **Peut-on battre le marché? Jensen (1968)**
- L'hypothèse de sur-réaction du marché De Bondt & Thaler (1985, 1987)

Étude de Jensen (1968) (2)

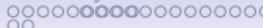
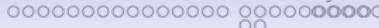
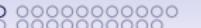
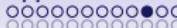
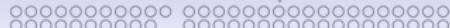
Première étude systématique sur les performances des fonds communs de placement.

Etude séparée des rendements annuels de 115 fonds entre 1945 et 1964. 115 régressions en séries temporelles, chaque régression étant la suivante:

$$R_{jt} - R_{ft} = \alpha_j + \beta_j(R_{mt} - R_{ft}) + u_{jt} \quad (18)$$

Variables :

- R_{jt} : rendement du portefeuille j à l'instant t .
- R_{ft} : rendement d'un actif sans risque (obligation d'État à 1 an).
- R_{mt} : rendement d'un portefeuille de marché.



Jensen's Alpha (α_j)

- α_j mesure si un fonds surperforme ou sous-performe l'indice de marché.
- β_j est proche de celui du MEDAF/CAPM.

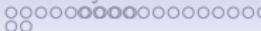
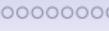
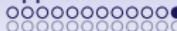
Élément	Moyenne	Médiane	Min.	Max.
$\hat{\alpha}$	-0,011	-0,009	-0,080	0,058
$\hat{\beta}$	0,840	0,848	0,219	1,405
N	17	19	10	20

Table: Statistiques résumées des résultats de régression (Jensen, 1968).

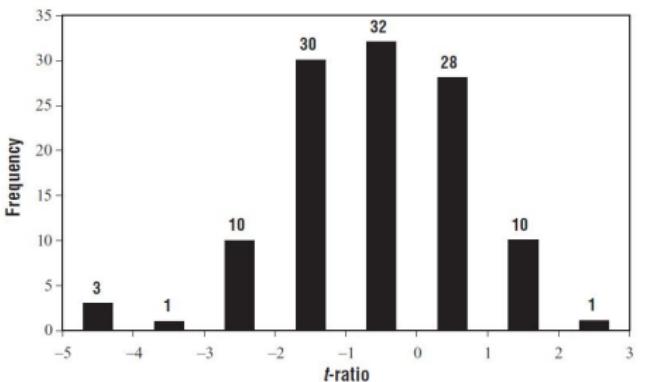
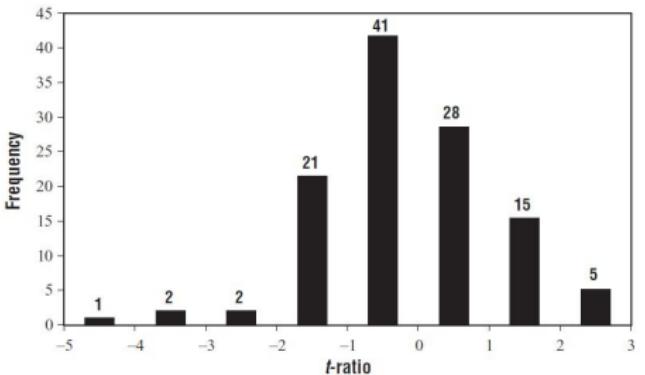
Statistiques résumées

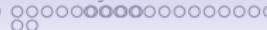
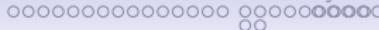
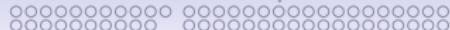
Résultats principaux :

- La majorité des fonds n'ont pas "battu le marché" (alpha moyen et médian négatifs).
- Quelques fonds ont généré des rendements anormaux significatifs (meilleur alpha : 0,058).
- Beta moyen : environ 0,85 (fonds moins risqués que le marché).
- Raisons possibles : investissement dans des actions blue chips (grandes valorisations, cours stables) plutôt que dans des small caps.



Approche graphique





Analyse des coefficients t de α_j

- Seuil critique : $|t| > 2,10$ (test bilatéral à 5%).
- Avant coûts de transaction :
 - 5 fonds surperforment significativement.
 - 5 fonds sous-performent significativement.
- Après coûts de transaction :
 - 1 fonds surperforme significativement.
 - 14 fonds sous-performent significativement.

Et sur une période plus récente?

Performance des fonds au Royaume-Uni: 76 fonds analysés mensuellement, 1979–2000, soit 257 observations par fonds.

	Moyenne	Min.	Max.	Médiane
Rendement mensuel moyen	1,0%	0,6%	1,4%	1,0%
Écart-type des rendements	5,1%	4,3%	6,9%	5,0%

Table: Statistiques résumées des rendements des fonds communs de placement (1979–2000).

Résultats

	Moy.	Min.	Max.	Médiane
α (%)	-0,02	-0,54	0,33	-0,03
β	0,91	0,56	1,09	0,91
Rapport t sur α	-0,07	-2,44	3,11	-0,25

- Beta moyen inférieur à 1 (fonds moins risqués que le marché).
- Avant coûts de transaction :
 - 9 fonds surperform significativement (α_j positif).
 - 7 fonds sous-perform significativement (α_j négatif).
- Après coûts de transaction : Fonds moyen incapable de générer des rendements excédentaires.

Résumé de Jensen (1968)

Résultats cohérents avec l'hypothèse d'efficience informationnelle des marchés financiers :

- Peu de fonds parviennent à générer des rendements anormaux positifs.
- Étude pionnière mais critiquée (taille limitée des échantillons pour les régressions).
- Implications pour l'évaluation de la performance des gestionnaires de fonds.

Applications

De multiples applications en économie et en finance existent:

- CAPM
- Peut-on battre le marché? Jensen (1968)
- L'hypothèse de sur-réaction du marché De Bondt & Thaler (1985, 1987)

Contexte de l'hypothèse de surréaction (3)

Études de DeBondt et Thaler (1985, 1987) :

- Les actions ayant mal performé sur 3 à 5 ans tendent à surperformer par la suite.
- Les perdants deviennent gagnants et vice versa.

Étude de Clare et Thomas (1995) :

- Analyse des rendements mensuels au Royaume-Uni (1955–1990).
- Données de toutes les entreprises cotées à la Bourse de Londres.

Explications possibles des surréactions

(1) Effet de taille :

- Les perdants sont souvent de petites entreprises, qui surperforment ensuite les grandes entreprises.
- La prise en compte de la taille réduit les rendements des perdants (Zarowin, 1990).

(2) Changements dans les rendements requis à l'équilibre :

- Les perdants ont des bêta CAPM plus élevés, reflétant un risque perçu plus élevé. Les bêta des perdants sont significativement plus élevés que ceux des gagnants (Ball & Kothari, 1989)

Méthodologie de Clare et Thomas (1995)

Échantillon aléatoire de 1 000 entreprises.

Calcul des rendements excédentaires mensuels (U_{it}) sur des périodes de 12, 24 ou 36 mois :

$$U_{it} = R_{it} - R_{mt}, \quad t = 1, \dots, n; \quad i = 1, \dots, 1000; \quad n = 12, 24, 36. \quad (19)$$

- R_{it} : rendement de l'action i .
- R_{mt} : rendement du marché.

Formation des portefeuilles

Calcul du rendement moyen mensuel pour chaque action i sur les périodes de 12, 24 ou 36 mois :

$$\bar{R}_i = \frac{1}{n} \sum_{t=1}^n U_{it} \quad (20)$$

- Classement des actions du rendement moyen le plus élevé au plus faible. Même durée n utilisée pour surveiller la performance des portefeuilles.

$$\bar{R}_D^t = \bar{R}_p^L - \bar{R}_p^W \quad (21)$$

- \bar{R}_p^L : rendement moyen mensuel des perdants (bottom 20%)
- \bar{R}_p^W : rendement moyen mensuel des gagnants (top 20%)
- \bar{R}_D^t : différence des rendements moyens mensuels.

Régressions des rendements excédentaires

Régression sur une constante seulement, avec :

$$\bar{R}_D = \alpha_1 + \eta_t \quad (22)$$

Régression incluant le risque de marché :

$$\bar{R}_D = \alpha_2 + \beta(R_{mt} - R_{ft}) + \eta_t \quad (23)$$

Résultats des coefficients (α_1 , α_2 et β) montrent que :

- α_1 est significatif pour 24 et 36 mois.
- Les différences de risque (β) ne sont pas significatives.

Tableau : Effet de surréaction sur le marché britannique

Période (n)	Bott.	Top	Diff.	Coeff.
Panel A : Tous les mois				
12 mois	0,0033	0,0036	-0,37%	$\hat{\alpha}_1 = -0,00031$ (0,29)
24 mois	0,0011	-0,0003	1,68%	$\hat{\alpha}_1 = 0,0014^{**}$ (2,01)
36 mois	0,0129	0,0115	1,56%	$\hat{\alpha}_1 = 0,0013$ (1,55)
Panel B : Hors janvier				
12 mois	–	–	–	$\hat{\alpha}_1 = -0,0007$ (-0,72)
24 mois	–	–	–	$\hat{\alpha}_1 = 0,0012^*$ (1,63)
36 mois	–	–	–	$\hat{\alpha}_1 = 0,0009$ (1,05)

Table: Résumé des rendements et coefficients pour l'effet de surréaction.

* Significatif à 10%, ** à 5%.

Conclusion sur les résultats

- Les perdants deviennent gagnants après 24 à 36 mois.
- Une partie de l'effet de surréaction est attribuée au mois de janvier.
- La taille des entreprises joue un rôle clé dans l'effet de surréaction.
- L'effet n'est pas expliqué par les différences de beta CAPM.

Les résultats semblent en contradiction avec l'hypothèse des marchés efficients. Implications pour les stratégies d'investissement basées sur les anomalies comportementales.

La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. ROC - Application à Dexia
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire

Étape (1/7) : Le modèle de régression multiple (modèle vrai)

Le modèle vrai de la population est le suivant :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + e \quad (24)$$

où :

- Y est la variable dépendante.
- X_1 est la variable explicative principale.
- X_2 est la variable explicative manquante.
- e est le terme d'erreur.

Dans ce modèle, b_1 mesure l'effet de X_1 sur Y , en contrôlant pour X_2 .

Étape (2/7) : Le modèle sous-spécifié (modèle estimé)

Supposons que X_2 ne soit pas observé ou disponible, et que nous estimions un modèle sans X_2 , ce qui donne :

$$\tilde{Y} = \tilde{b}_0 + \tilde{b}_1 X_1 \quad (25)$$

où \tilde{b}_1 est l'estimateur de b_1 à partir du modèle sous-spécifié.

Étape (3/7) : Biais causé par l'omission de X_2

Le biais dû à l'omission de X_2 se produit parce que X_2 affecte Y dans le modèle vrai, mais est ignoré dans le modèle estimé. Pour dériver explicitement ce biais, nous devons considérer la relation entre \tilde{b}_1 et les paramètres du modèle vrai.

On peut montrer que :

$$\tilde{b}_1 = b_1 + b_2 \tilde{\delta}_1 \quad (26)$$

où $\tilde{\delta}_1$ est la pente de la régression simple de X_2 sur X_1 , c'est-à-dire :

$$\tilde{\delta}_1 = \frac{\sum_{i=1}^N (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2)}{\sum_{i=1}^n (X_{i1} - \bar{X}_1)^2} = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} \quad (27)$$

Cela montre que \tilde{b}_1 , l'estimateur obtenu en omettant X_2 , est biaisé si $b_2 \neq 0$ et $\tilde{\delta}_1 \neq 0$.

Étape (4/7) : Espérance de \tilde{b}_1

L'espérance de \tilde{b}_1 , en conditionnant sur X_1 et X_2 , est donnée par :

$$E(\tilde{b}_1) = E(b_1 + b_2 \tilde{\delta}_1) \quad (28)$$

Comme b_1 et b_2 sont des constantes, cela donne :

$$E(\tilde{b}_1) = b_1 + b_2 \tilde{\delta}_1 \quad (29)$$

Le biais de \tilde{b}_1 est donc :

$$\text{Biais}(\tilde{b}_1) = E(\tilde{b}_1) - b_1 = b_2 \tilde{\delta}_1 \quad (30)$$

Étape (5/7) : Interprétation du biais

Le biais est proportionnel à deux facteurs :

- b_2 : L'effet de X_2 sur Y .
- $\tilde{\delta}_1$: La corrélation entre X_1 et X_2 .

Cela signifie que :

- Si $b_2 > 0$ et X_1 et X_2 sont positivement corrélés ($\tilde{\delta}_1 > 0$), alors \tilde{b}_1 sera surestimé (biais positif).
- Si $b_2 > 0$ et X_1 et X_2 sont négativement corrélés ($\tilde{\delta}_1 < 0$), alors \tilde{b}_1 sera sous-estimé (biais négatif).

Étape (6/7) : Cas où \tilde{b}_1 est sans biais

Il existe deux cas où \tilde{b}_1 sera sans biais :

- $b_2 = 0$: Si X_2 n'a aucun effet sur Y , alors il n'y a pas de biais dû à l'omission de X_2 .
- $\tilde{\delta}_1 = 0$: Si X_1 et X_2 sont non corrélés, l'omission de X_2 n'introduit pas de biais.

Étape (7/7) : Conséquences pratiques

En pratique, il est souvent difficile de savoir si une variable comme X_2 est omise. Cependant, on peut faire des hypothèses raisonnables sur :

- Le signe de b_2 (l'effet de X_2 sur Y).
- Le signe de $\tilde{\delta}_1$ (la corrélation entre X_1 et X_2).

	$\text{Corr}(X_1, X_2) > 0$	$\text{Corr}(X_1, X_2) < 0$
$b_2 > 0$	Biais positif	Biais négatif
$b_2 < 0$	Biais négatif	Biais positif

Variance des estimateurs des MCO

Sous les hypothèses **H1** à **H5**, conditionnées aux valeurs d'échantillon des variables explicatives, la variance de \hat{b}_j est donnée par :

$$\text{Var}(\hat{b}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad j = 1, 2, \dots, K \quad (31)$$

où :

- $SST_j = \sum_{i=1}^N (X_{ij} - \bar{X}_j)^2$ est la variation totale dans X_j dans l'échantillon.
- R_j^2 est le R^2 de la régression de X_j sur les autres variables explicatives.

Composantes de la variance des MCO

La variance de \hat{b}_j dépend de trois facteurs :

- 1. La variance de l'erreur σ^2** : Plus σ^2 est grande, plus la variance des estimateurs des MCO est grande.
- 2. La variation totale dans X_j (SST_j)** : Plus la variation de X_j est grande, plus la variance de \hat{b}_j est petite. Pour une estimation précise, il est préférable d'avoir une grande variation de l'échantillon dans X_j .
- 3. La multicolinéarité (R_j^2)** : Si X_j est fortement corrélé avec d'autres variables explicatives, alors R_j^2 sera proche de 1, ce qui augmente la variance de \hat{b}_j . Une forte relation linéaire entre X_j et les autres variables explicatives conduit à une variance élevée pour les estimateurs des MCO.

Multicolinéarité et variance

Le terme R_j^2 provient de la régression de X_j sur toutes les autres variables explicatives. Si R_j^2 est proche de 1, cela signifie que X_j est fortement corrélé avec les autres variables explicatives, entraînant une augmentation de $\text{Var}(\hat{b}_j)$.

Lorsque $R_j^2 = 1$, cela signifie que X_j est une combinaison linéaire parfaite des autres variables, ce qui est exclu par l'hypothèse **H3**. En revanche, lorsque R_j^2 est proche de zéro, cela signifie que X_j est faiblement corrélé avec les autres variables, ce qui minimise la variance de \hat{b}_j .

Multicolinéarité et variance

Exemple de multicolinéarité: taux d'approbation des crédits immobiliers (Y) et quartier.

Nous sommes tentés d'ajouter un certain nombre de variables plus ou moins reliés à de la discrimination:

- Revenu moyen, âge moyen de la population, nombre de guichets bancaires, chômage, score des élections de tel ou tel parti...
- Ces variables risquent d'être hautement corrélées. Donc variance des \hat{b}_j en augmentation. Difficultés à établir des relations causales.

→ Comment choisir en fonction de la sévérité de la multicolinéarité? Indice VIF

Multicolinéarité et variance

$$\text{Var}(\hat{b}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)} = \frac{\sigma^2}{SST_j} VIF_j, \quad j = 1, 2, \dots, K \quad (32)$$

Variance inflation factor (VIF) égal à $\frac{1}{1-R_j}$.

Seuil d'alerte pour chaque variable explicative: 10, voir dans certains cas 5...

Variances dans les modèles mal spécifiés

Le choix d'inclure ou non une variable dans un modèle de régression peut être analysé à travers le compromis entre biais et variance.

Nous avons montré dans l'espérance des estimateurs le biais induit par l'omission d'une variable pertinente dans un modèle où le modèle vrai contient deux variables explicatives.

Nous allons continuer cette analyse en comparant les variances des estimateurs des MCO.

Modèle vrai de population

Le modèle de population vrai, qui satisfait les hypothèses de Gauss-Markov, est le suivant :

$$Y = b_0 + b_1 X_1 + b_2 X_2 + e \quad (33)$$

Nous considérons deux estimateurs de b_1 :

- \hat{b}_1 : L'estimateur provenant de la régression multiple.
- \tilde{b}_1 : L'estimateur provenant de la régression simple, en omettant X_2 .

Estimations multiples vs simples

L'estimateur \hat{b}_1 provient de la régression multiple suivante :

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 \quad (34)$$

Autrement dit, nous incluons X_2 dans le modèle avec X_1 . En revanche, l'estimateur \tilde{b}_1 provient de la régression simple suivante :

$$\tilde{Y} = \tilde{b}_0 + \tilde{b}_1 X_1 \quad (35)$$

Lorsque $b_2 \neq 0$, cette équation exclut une variable pertinente du modèle, ce qui induit un biais dans \tilde{b}_1 , sauf si X_1 et X_2 sont non corrélés. En revanche, \hat{b}_1 est non biaisé pour b_1 , quelle que soit la valeur de b_2 .

Comparaison des variances :

La conclusion que \hat{b}_1 est toujours préférable à \tilde{b}_1 ne tient pas lorsque l'on prend en compte la variance. En conditionnant sur les valeurs de X_1 et X_2 dans l'échantillon, nous avons :

$$\text{Var}(\hat{b}_1) = \frac{\sigma^2}{SST_1(1 - R_1^2)} \quad (36)$$

où SST_1 est la variation totale de X_1 , et R_1^2 est le R^2 de la régression de X_1 sur X_2 . De plus, une modification simple de la preuve en Chapitre 2 pour la régression à deux variables montre que

$$\text{Var}(\tilde{b}_1) = \frac{\sigma^2}{SST_1} \quad (37)$$

La comparaison des 2 équations montre que $\text{Var}(\tilde{b}_1)$ est toujours plus petite que $\text{Var}(\hat{b}_1)$, sauf si X_1 et X_2 sont non corrélés, auquel cas les deux estimateurs \tilde{b}_1 et \hat{b}_1 sont les mêmes.

Samuel Ligonièvre

Boîte à outils

Il y a deux cas:

1. Lorsque $b_2 \neq 0$, \tilde{b}_1 est biaisé, \hat{b}_1 est non biaisé, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.
2. Lorsque $b_2 = 0$, \tilde{b}_1 et \hat{b}_1 sont tous les deux non biaisés, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.

Quand faut-il préférer \tilde{b}_1 ou \hat{b}_1 ? Cela dépend de (1) des différences de R^2 , (2) de l'importance du risque de multicolinéarité, et enfin (3) de l'importance du biais potentiel.

Boîte à outils

1. Lorsque $b_2 \neq 0$, \tilde{b}_1 est biaisé, \hat{b}_1 est non biaisé, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.
2. Lorsque $b_2 = 0$, \tilde{b}_1 et \hat{b}_1 sont tous les deux non biaisés, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.

Dans ce cas, les choses sont simples: il est clair que \tilde{b}_1 est préférable si $b_2 = 0$, si X_2 n'a pas d'effet partiel sur Y . Ajouter la deuxième variable ne va qu'exacerber le problème de multicolinéarité et donc b_1 est un estimateur moins efficient. En d'autres termes, pas de biais mais une variance plus grande de l'estimateur.

Boîte à outils

1. Lorsque $b_2 \neq 0$, \tilde{b}_1 est biaisé, \hat{b}_1 est non biaisé, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.
2. Lorsque $b_2 = 0$, \tilde{b}_1 et \hat{b}_1 sont tous les deux non biaisés, et $\text{Var}(\tilde{b}_1) < \text{Var}(\hat{b}_1)$.

Ce cas est un plus difficile: Le cas où $b_2 \neq 0$ est plus difficile.

Omettre X_2 du modèle conduit à un estimateur biaisé de b_1 .

Traditionnellement, les économètres suggèrent de comparer la taille probable du biais dû à l'omission de X_2 avec la réduction de la variance (résumée dans R^2_1) pour décider si X_2 doit être inclus.

Cas où $b_2 \neq 0$:

Lorsque $b_2 \neq 0$, il y a deux raisons favorables à l'inclusion de X_2 dans le modèle :

- Le biais de \tilde{b}_1 ne diminue pas à mesure que la taille de l'échantillon augmente. En revanche, $\text{Var}(\tilde{b}_1)$ et $\text{Var}(\hat{b}_1)$ diminuent toutes les deux à mesure que N devient grand, ce qui signifie que la multicolinéarité induite par l'ajout de X_2 devient moins importante avec un grand échantillon. Dans les grands échantillons, nous préférerions \hat{b}_1 .
- La variance de \tilde{b}_1 , conditionnée uniquement sur X_1 , est plus grande que celle présentée dans initialement. En effet, lorsque $b_2 \neq 0$ et que X_2 est exclu du modèle, la variance de l'erreur augmente car l'erreur contient en fait une partie de X_2 .

La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. **Variables qualitatives - Application aux crises bancaires**
6. ROC - Application à Dexia
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire

Variables qualitatives

Usage de variables qualitatives (dummies ou variables catégorielles):

- Usage à gauche: modèle de probabilité linéaire, logit, probit
- Usage à droite: interprétation particulière

Cas du Y en dummies On cherche à expliquer les valeurs de Y grâce à X , c'est-à-dire à estimer la probabilité que $Y_i = 1$ sachant X_i (ou que $Y_i = 0$, ce qui revient au même).

Usage des MCO

Le modèle de probabilité linéaire est une méthode simple pour estimer la relation entre une variable binaire Y et des variables explicatives X . Cependant, il présente des avantages et des inconvénients :

- **Avantages :**
 - Facilité de mise en œuvre et d'interprétation des coefficients.
 - Modèle linéaire classique qui peut être estimé par les moindres carrés ordinaires (OLS).
- **Inconvénients :**
 - Les prédictions de probabilités peuvent sortir de l'intervalle.
 - Hypothèse irréaliste de linéarité entre X et $P(Y = 1|X)$.
 - Problèmes de variance hétéroscléastique.

Logit et probit

$$P(Y_i|X_i) = P(X_i\theta + e_i \geq 0|X_i) = P(X_i\theta \geq -e_i|X_i) = F_{-e}(X_i\theta) \quad (38)$$

Interprétation via le signe de l'effet marginal pour la variable j :

$$\frac{\partial F(X_i\theta)}{\partial X_{ij}} = f(X_i\theta)\theta_j \quad (39)$$

Dans ces deux cas, la loi des résidus est symétrique, on peut donc remplacer F_{-e} par F_e .

La seule différence entre les modèles probit et logit est la spécification de F . Dans un Probit, F est une fonction de répartition d'une gaussienne centrée réduite. Dans un Logit, F correspond à la loi logistique.

Dependent variable

Multiple possibilities:

- Dummy variable: recession, depression, peak, trough, banking crisis, financial crisis
- Continuous variable: growth rate of a variable. GDP: in current prices, in constant prices? GDP or GDP per capita?
- Cycle characteristics: duration, amplitude
- Business cycles vs financial cycles?

Not the same econometric model!

Variable dépendante

Multiples possibilités :

- Variable binaire : récession, dépression, pic, creux, crise bancaire, crise financière
- Variable continue : taux de croissance d'une variable. PIB : en prix courants, en prix constants ? PIB ou PIB par habitant ?
- Caractéristiques du cycle : durée, amplitude
- Cycles économiques vs cycles financiers ?

Modèle économétrique différent !

Cas des crises bancaires

Sur la base de la méthodologie de Laeven et Valencia (FMI, 2008, 2013, 2020), plusieurs conditions sont requises pour définir une crise bancaire systémique :

1. Signes significatifs de détresse financière dans le système bancaire (indiqués par des retraits bancaires massifs, des pertes importantes dans le système bancaire et/ou des liquidations bancaires).
2. Mesures significatives d'intervention en matière de politique bancaire en réaction à des pertes importantes dans le système bancaire.

Ils considèrent que la première année où ces deux critères sont remplis correspond à l'année où la crise devient systémique. Cela permet de dater la crise aux premiers signes de problèmes majeurs dans le système bancaire.

Samuel Ligonière

Cas des crises bancaires

Première condition Lorsque les pertes dans le secteur bancaire et/ou les liquidations sont sévères, le premier critère est considéré comme une condition suffisante pour dater une crise bancaire systémique. Cette définition est opérationnalisée en considérant que les pertes sont sévères lorsque :

1. le système bancaire d'un pays enregistre des pertes significatives, avec un ratio de créances douteuses supérieur à 20
2. les coûts de restructuration budgétaire du secteur bancaire sont suffisamment élevés, dépassant 5

Cas des crises bancaires

Deuxième condition Se fier exclusivement au premier critère est problématique, car il n'est pas toujours simple de quantifier le degré de détresse financière dans un système bancaire, en particulier dans les pays à revenu faible et intermédiaire. De plus, les pertes peuvent être atténuées par des réponses politiques. Pour résoudre ce problème, nous nous appuyons également sur le second critère si l'intervention politique remplit l'exigence d'être significative.

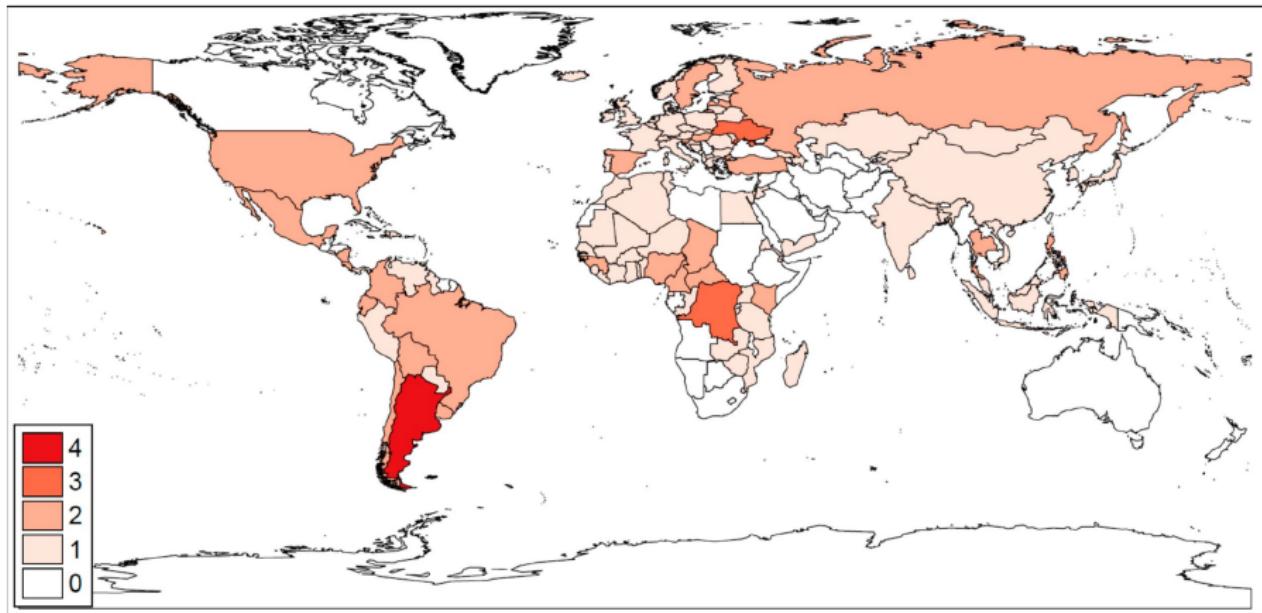
Cas des crises bancaires

Deuxième condition Nous considérons qu'une intervention politique dans le secteur bancaire est significative si au moins trois des six mesures suivantes ont été mises en place :

1. gel des dépôts et/ou fermetures temporaires de banques ;
2. nationalisations bancaires significatives ;
3. coûts budgétaires de restructuration bancaire (au moins 3)
4. soutien massif à la liquidité (au moins 5)
5. garanties importantes mises en place ;
6. rachats d'actifs significatifs (au moins 5)

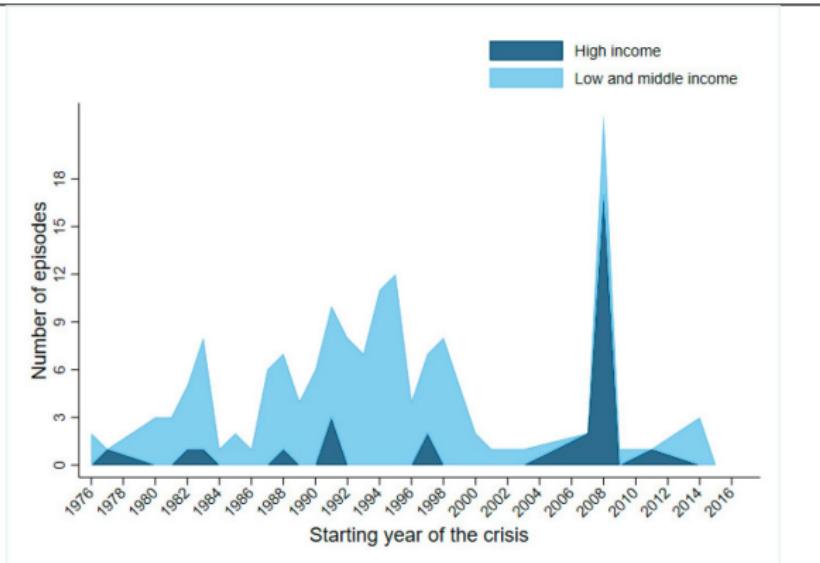
Cas des crises bancaires

Figure 1. Frequency of Systemic Banking Crises Around the World, 1970–2017



Cas des crises bancaires

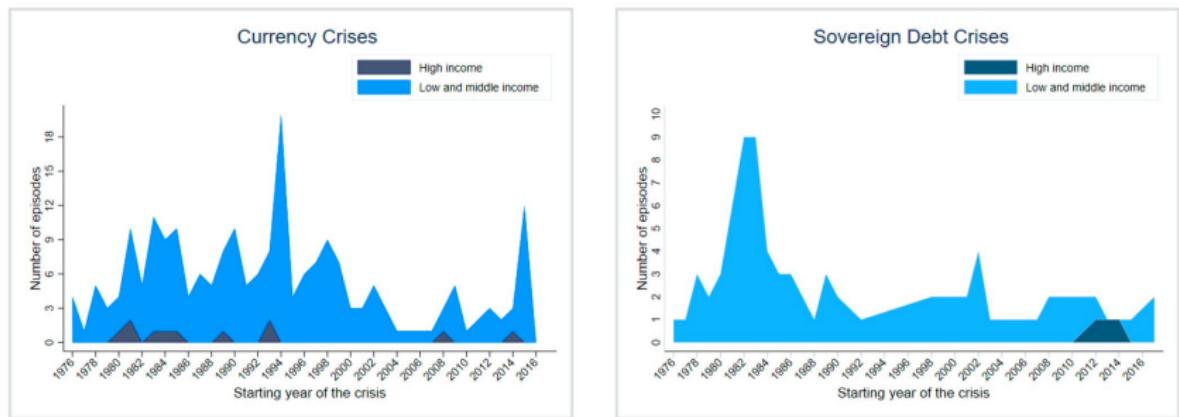
Figure 2. Systemic Banking Crises Episodes by Income Level 1970–2017



Source: Authors' calculations.

Cas des crises bancaires

Figure 3. Currency and Sovereign Debt Crises Episodes by Income level

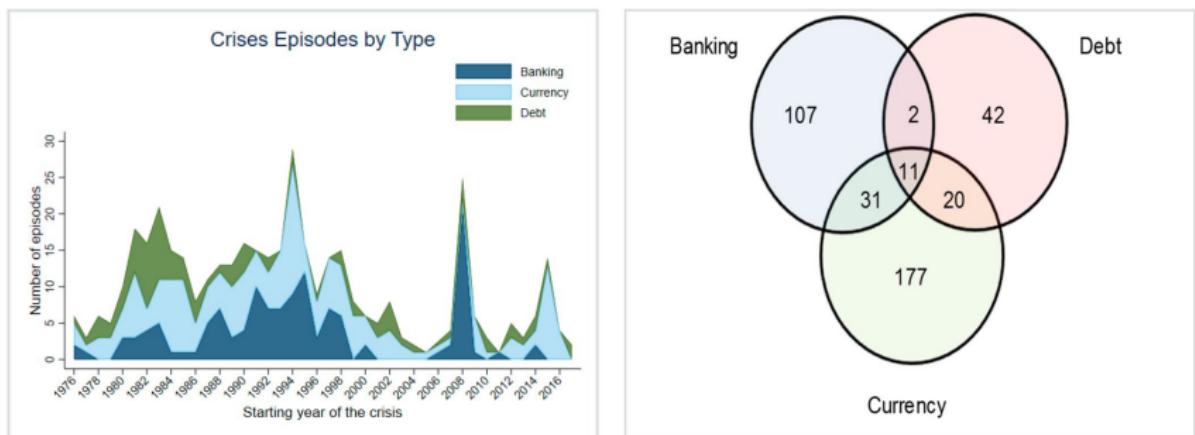


Source: Authors' calculations.

Note: We distinguish high income countries from low and middle-income countries, based on the level of GNI per capita in US\$ in the year in which the banking crisis episode started. The classification is assigned by comparing such level of GNI per capita to the income thresholds defined by the World Bank for that same year.

Cas des crises bancaires

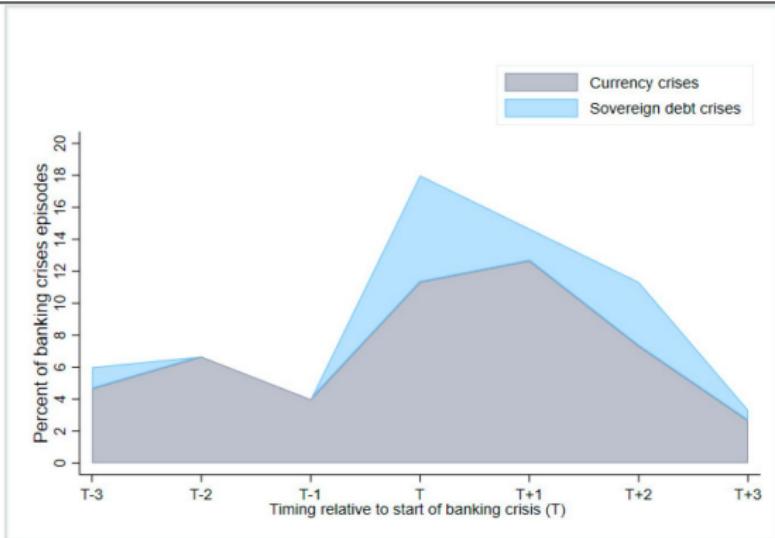
Figure 4. Financial Crises by Type



Source: Authors' calculations.

Cas des crises bancaires

Figure 5. Sequencing of Crises



Source: Authors' calculations.

Note: The figure is constructed by selecting banking crises episodes and plotting the percentage of them that were followed, coincided, or were preceded by a sovereign or currency crisis, with T denoting the start of the banking crisis.

La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. **ROC - Application à Dexia**
7. Filtres temporels
8. Vers le NLP - Application à la politique monétaire

Introduction à la courbe ROC

Courbe ROC (Receiver Operating Characteristic) :

- Une représentation graphique des performances d'un modèle de classification binaire.
- Trace le **Taux de Vrais Positifs (TPR)** contre le **Taux de Faux Positifs (FPR)** pour différentes valeurs de seuil.

Pourquoi utiliser les courbes ROC ?

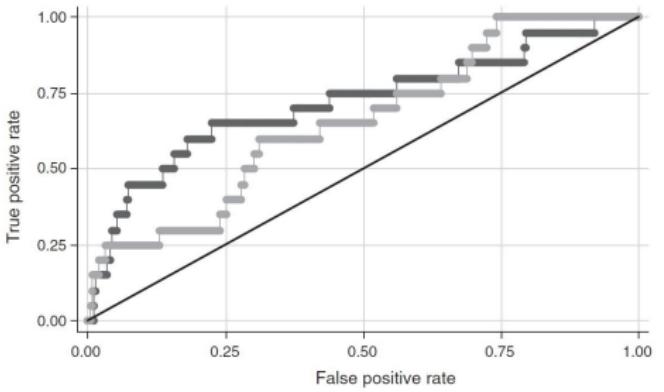
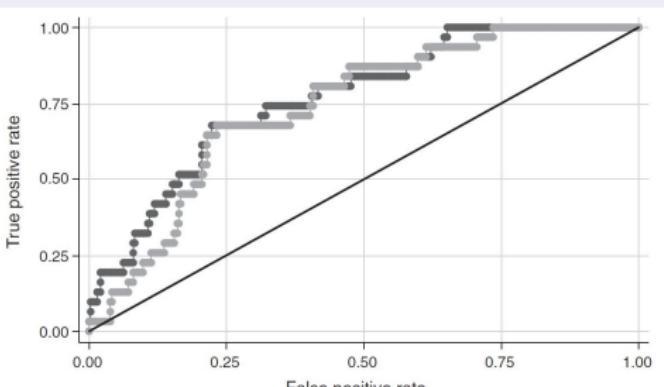
- Évaluer le compromis entre sensibilité (TPR) et spécificité (1 - FPR).
- Comparer différents classifieurs indépendamment du seuil de décision.

Métriques clés

Taux de Vrais Positifs (TPR) et Taux de Faux Positifs (FPR)

Aire sous la courbe (AUROC) :

- Mesure la performance globale du classifieur.
- AUC = Probabilité que le classifieur classe une instance positive aléatoire plus haut qu'une instance négative aléatoire.
- Va de 0,5 (choix aléatoire) à 1 (classification parfaite).



Courbe ROC

- Chaque point sur la courbe ROC correspond à un seuil spécifique.
- La ligne diagonale représente un classifieur aléatoire.
- Plus la courbe est proche du coin supérieur gauche, meilleur est le classifieur.

Comment calculer l'AUROC ? Étapes :

1. Trier les prédictions par probabilité.
2. Calculer le TPR et le FPR pour chaque seuil.
3. Tracer le TPR contre le FPR.
4. Calculer l'aire sous la courbe ROC à l'aide d'une intégration numérique (par exemple, méthode des trapèzes).

Avantages et limitations

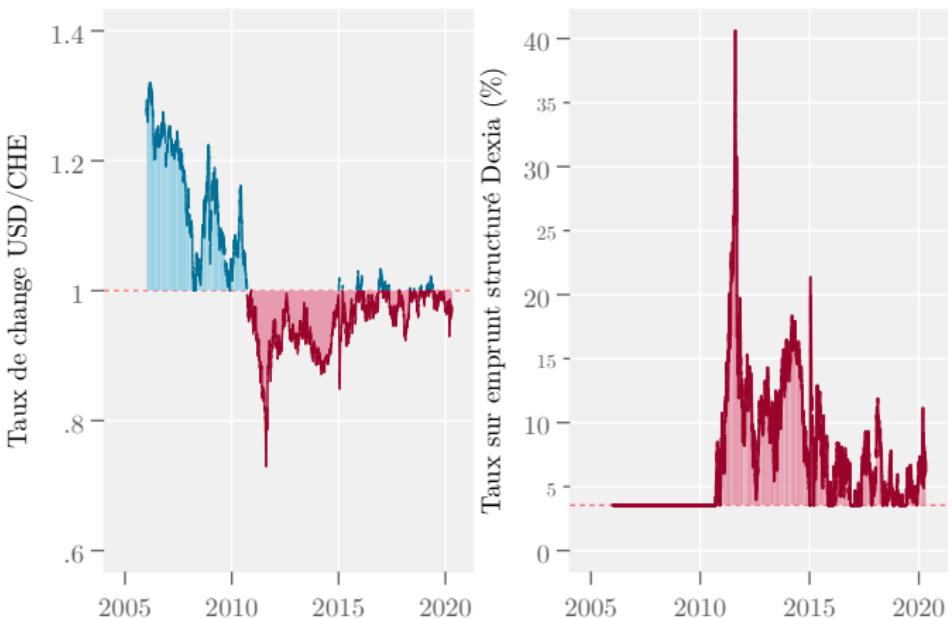
Avantages :

- Fournit une évaluation complète des performances du modèle.
- Analyse indépendante du seuil.

Limitations :

- Ne tient pas compte directement du déséquilibre des classes.
- Les valeurs d'AUC peuvent être trompeuses pour des ensembles de données fortement déséquilibrés.

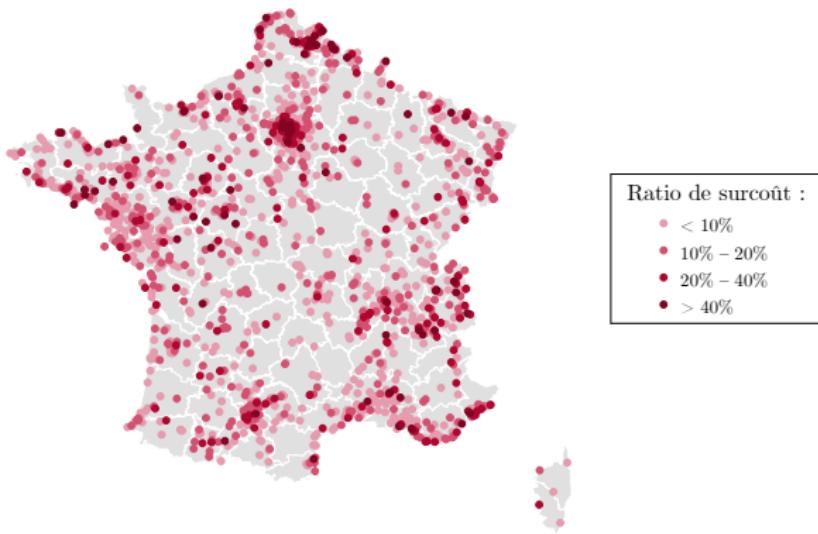
Application à la crise Dexia



Face à l'appréciation du franc suisse, instauration du taux plancher
 $1.2CF = 1\text{ euro}$ puis suppression en 2015.

Samuel Ligonière

Application à la crise Dexia



Cas des produits structurés - 1500 communes françaises

Samuel Ligonière

(Année universitaire 2024-2025)

M2 Advanced Econometrics

131 / 185

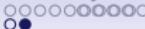
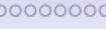
Table: Critères réglementaires.

Régressions Logit	(1)	(2)	(3)	(4)	(5)	(6)	
Autres dimensions comme variables de contrôle:	non	non	non	non	non	oui	Effets margin.
(a) Dette par habitant	0.063* (0.028)	0.064* (0.028)	0.063* (0.028)	0.074* (0.033)	0.044 (0.025)	0.023 (0.035)	0.001 (0.002)
(b) Part Dexia dans l'endettement total		0.050 (0.026)	0.051 (0.026)	0.050 (0.026)	0.024 (0.028)	0.110** (0.039)	0.006** (0.002)
(c) Capacité de désendettement			0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	-0.002 (0.002)	-0.000 (0.000)
(d) Coefficient de mobilisation du potentiel fiscal				-0.342 (0.285)	-0.370 (0.340)	-0.198 (0.395)	-0.011 (0.022)
Ratio de surcoût total					10.308*** (1.162)	9.614*** (1.010)	0.540*** (0.061)
Observations	1542	1542	1542	1542	1542	1542	-
Auroc	0.687	0.749	0.741	0.629	0.858	0.872	-
Samuel Ligonnière	(0.023)	(0.019)	(0.019)	(0.025)	(0.018)	(0.017)	-

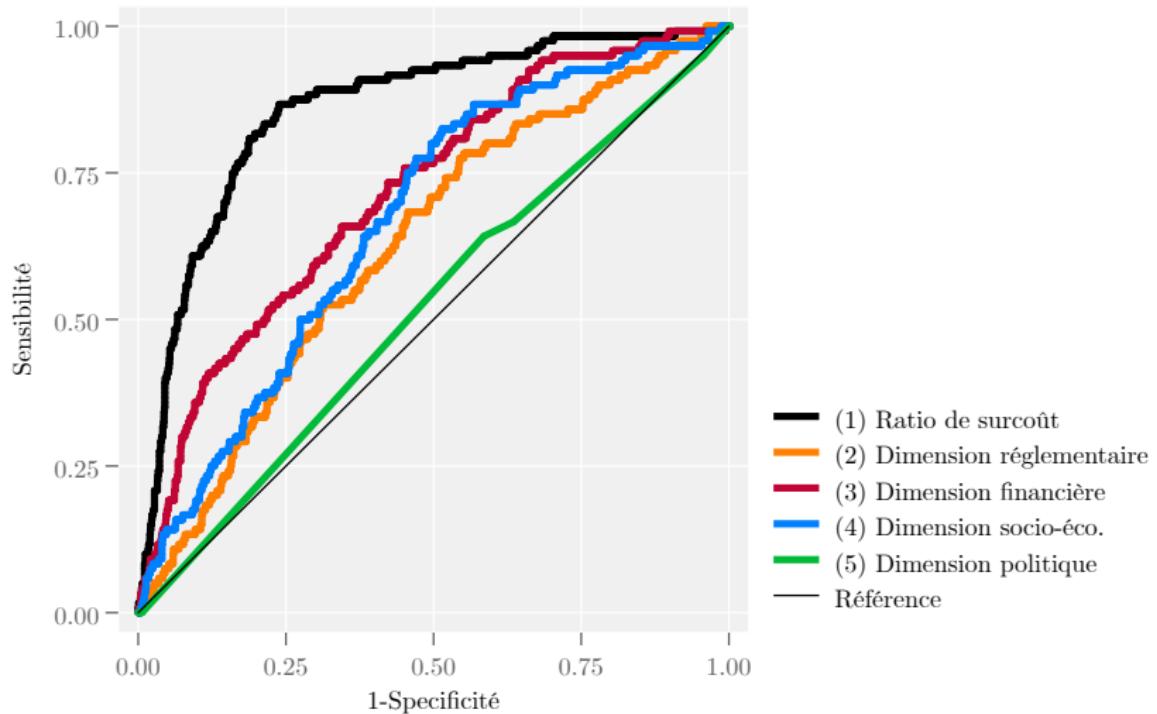
Table: Autres dimensions financières.

Régressions Logit	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Autres dimensions comme variables de contrôle:	non	non	non	non	non	non	oui	Effets margin.
(a) Taux d'équipement	-3.228*** (0.906)	-3.289*** (0.868)	-3.350*** (0.887)	-2.227** (0.838)	-2.008* (0.856)	-2.386* (1.045)	-2.704* (1.279)	-0.152* (0.072)
(b) Taux d'endettement		0.711*** (0.157)	0.673*** (0.165)	0.831*** (0.180)	0.752*** (0.193)	0.585** (0.190)	0.819*** (0.237)	0.046*** (0.014)
(c) Taux d'autofinancement des nouveaux investissements			-0.180 (0.101)	0.204 (0.114)	0.184 (0.125)	0.051 (0.142)	0.124 (0.151)	0.007 (0.008)
(d) Taux de capacité d'autofinancement				-6.775*** (1.474)	-6.061*** (1.461)	-5.055** (1.642)	-4.209* (1.894)	-0.236* (0.106)
(e) Taux de rigidité des charges de structure					3.404* (1.662)	2.772 (1.699)	1.563 (1.923)	0.088 (0.107)
Ratio de surcoût total						9.840*** (1.032)	9.614*** (1.010)	0.540*** (0.061)
Observations	1542	1542	1542	1542	1542	1542	1542	—
Auroc	0.621 (0.024)	0.678 (0.024)	0.687 (0.024)	0.715 (0.023)	0.716 (0.024)	0.861 (0.017)	0.872 (0.017)	—

Samuel Ligonièvre



ROC Dexia



La boîte à outils économétriques

1. Théorème Gauss-Markov
2. Tests et p-value
3. Applications
4. Choix du modèle: biais, variance des estimateurs et multicolinéarité
5. Variables qualitatives - Application aux crises bancaires
6. ROC - Application à Dexia
7. **Filtres temporels**
8. Vers le NLP - Application à la politique monétaire

Why date cycles?

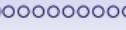
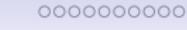
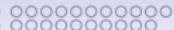
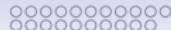
Economic and financial cycles influence fiscal and monetary policy.

Researchers need reference chronologies to validate models.

International comparisons rely on consistent dating frameworks.

Examples:

- NBER (US): Business Cycle Dating Committee (since 1978).
- CEPR (EU): Euro Area Business Cycle Dating Committee (since 2003).
- National initiatives in Brazil, France, Spain, and Canada.



Cycles

At a minimum, three key aspects must be considered for any cycle:

1. How do I define it?
2. How do I date it?
3. If necessary, how do I filter the data to differentiate between the trend and short-term, medium-term, and long-term components?

Definition of the cycle

Initial work conducted at the NBER by Arthur Burns and Wesley Mitchell (1946) provided definitions still widely used today:

- No distinction between business cycles and cycles of economic activity: they focus on short-term dynamics.
- Cycles are defined as the succession of phases of economic contraction (negative growth, i.e., recessions) and phases of economic expansion (positive growth, i.e., expansions).
- Periods are bounded by peaks (the highest level of activity) and troughs (the lowest level of activity), corresponding to the turning points of a cycle.

Definition of the cycle

This definition highlights two key characteristics of economic cycles, which are essential for developing econometric approaches:

- 1. Co-movement:** Representative economic variables tend to move together during the cycle.
- 2. Distinct regimes:** Two distinct phases—expansion and recession—can differ in terms of duration, amplitude, and volatility.

Definition of the cycle

Two alternative concepts, though rarely used in practice:

1. **The output gap cycle:** Represents deviations from the long-term trend (Mintz, NBER 1969). This refers to the economy's potential but is difficult to estimate.
2. **The acceleration cycle:** Distinguishes phases of increasing and decreasing growth rates (i.e., acceleration and deceleration). The peak of this cycle represents the local maximum of the growth rate, and the trough represents the local minimum. However, this approach can be overly volatile, and GDP growth alone already provides significant information.

Definition of the cycle

Two alternative concepts, though rarely used in practice:

1. **The output gap cycle:** Represents deviations from the long-term trend (Mintz, NBER 1969). This refers to the economy's potential but is difficult to estimate.
2. **The acceleration cycle:** Distinguishes phases of increasing and decreasing growth rates (i.e., acceleration and deceleration). The peak of this cycle represents the local maximum of the growth rate, and the trough represents the local minimum. However, this approach can be overly volatile, and GDP growth alone already provides significant information.

Definition of the cycle: recessions

The simplest element to define is recessions.

1. **The simplest model:** The rule of two consecutive quarters of negative growth. However:
 - The 2001 US recession caused by the bursting of the internet bubble would not qualify as a recession.
 - This rule identifies the beginning but does not pinpoint the end of a recession.
2. **Burns and Mitchell (1946):** They characterize a recession as "a significant decline in economic activity across sectors, lasting more than a few months. This decline should normally be reflected in GDP, employment, industrial production, and sales in the manufacturing and trade sectors." → Harding, Pagan (2002)

Definition of the cycle: recessions

The simplest element to define is recessions.

1. **The simplest model:** The rule of two consecutive quarters of negative growth. However:
 - The 2001 US recession caused by the bursting of the internet bubble would not qualify as a recession.
 - This rule identifies the beginning but does not pinpoint the end of a recession.
2. **Burns and Mitchell (1946):** They characterize a recession as "a significant decline in economic activity across sectors, lasting more than a few months. This decline should normally be reflected in GDP, employment, industrial production, and sales in the manufacturing and trade sectors." → Harding, Pagan (2002)

Definition of the cycle: recessions

Historical recessions in France (1970-2019)

Key dates identified using the approach of Burns and Mitchell (1946) and Harding and Pagan (2002):

- 1974-75: First Oil Shock.
- 1980: Second Oil Shock.
- 1992-93: Investment Cycle Crisis.
- 2008-09: Global Financial Crisis.
- 2012-13: Eurozone Sovereign Debt Crisis.

Severity comparison:

- **Great Recession (2008-09):** The most severe, with GDP falling by 3.9
- **Oil Shocks:** Shorter duration but significant disruptions.

Definition of the cycle: recessions

In line with Burns and Mitchell (1946) and Harding and Pagan (2002), the "3D Rule":

1. **Duration:** Minimum of 6 months for a recession to be valid.
2. **Depth:** Not just a slightly negative GDP growth rate.
Conversely, a sharp decline over three months can also qualify as a recession.
3. **Diffusion:** The impact must affect multiple sectors of the economy and not be confined to one industry or geographic area pulling GDP down.

→ The Harding and Pagan (2003) algorithm characterizes these aspects (shapes of recessions: convex or concave?) and incorporates expert analysis.

Definition of the cycle: recessions

In line with Burns and Mitchell (1946) and Harding and Pagan (2002), the "3D Rule":

1. **Duration:** Minimum of 6 months for a recession to be valid.
2. **Depth:** Not just a slightly negative GDP growth rate.
Conversely, a sharp decline over three months can also qualify as a recession.
3. **Diffusion:** The impact must affect multiple sectors of the economy and not be confined to one industry or geographic area pulling GDP down.

→ The Harding and Pagan (2003) algorithm characterizes these aspects (shapes of recessions: convex or concave?) and incorporates expert analysis.

Definition of the cycle: recessions

In line with Burns and Mitchell (1946) and Harding and Pagan (2002), the "3D Rule":

1. **Duration:** Minimum of 6 months for a recession to be valid.
2. **Depth:** Not just a slightly negative GDP growth rate.
Conversely, a sharp decline over three months can also qualify as a recession.
3. **Diffusion:** The impact must affect multiple sectors of the economy and not be confined to one industry or geographic area pulling GDP down.

→ The Harding and Pagan (2003) algorithm characterizes these aspects (shapes of recessions: convex or concave?) and incorporates expert analysis.

Definition of the cycle

National Bureau of Economic Research (NBER)

The US Business Cycle Dating Committee was established in 1978.
Its methodology includes:

- Focus on broad aggregates like GDP, employment, and industrial production.
- Criteria: Duration, Depth, Diffusion ("3D Rule").
- Example: The "Great Recession" identified in 2008-2009.

Definition of the cycle

Euro Area Business Cycle Dating Committee (CEPR)

Founded in 2003 and modeled after the NBER, this committee focuses on dating business cycles in the Eurozone.

Notable insights:

- Synchronization of cycles across member states.
- The 2012-2013 Eurozone debt crisis and its impacts.

Peaks and troughs define transitions. Example: The COVID-19 recession starting in Q4 2019.

Cycles

At a minimum, three key aspects must be considered for any cycle:

1. How do I define it?
2. **How do I date it?**
3. If necessary, how do I filter the data to differentiate between the trend and short-term, medium-term, and long-term components?

BBQ Algorithm - Harding & Pagan (2002)

Objective: Identify turning points (peaks and troughs) in economic time series to date cycles.

Key Features:

- No explicit trend-cycle decomposition: no FILTERS!
- Relies on rules based on **direction**, **duration**, and **depth**.
- Designed for identifying business cycle phases in raw or seasonally adjusted data.

Key Parameters of BBQ

The algorithm uses three main parameters:

- **Window (w)**: Size of the neighborhood used to detect local extremes → filters noise and detects local maxima/minima.
- **Phase (p)**: Minimum duration between a peak and a trough → ensures realistic separation between turning points.
- **Cycle (c)**: Minimum duration of a complete cycle → filters out fluctuations that are too short to qualify as cycles.

Detecting Local Extremes (ltpoint)

Goal: Identify local maxima (peaks) and minima (troughs) within a defined window w .

Mathematical Rule:

$$\text{ltpoint}(t) = \begin{cases} +1 & \text{if } x_t > x_{t-1} \text{ and } x_t > x_{t+1} \quad (\text{Peak}) \\ -1 & \text{if } x_t < x_{t-1} \text{ and } x_t < x_{t+1} \quad (\text{Trough}) \\ 0 & \text{otherwise.} \end{cases}$$

Example: For a series $x_t = \{100, 105, 102, 98, 110\}$:

- $x_2 = 105$ is a peak (+1).
- $x_4 = 98$ is a trough (-1).

Adding the Second Turning Point

Purpose: Identify the second turning point (peak or trough) after the first one, respecting the minimum phase duration (p).

Rules:

- After a **peak**, look for a **trough** such that:

$$t_{\text{trough}} - t_{\text{peak}} \geq p$$

- After a **trough**, look for a **peak** such that:

$$t_{\text{peak}} - t_{\text{trough}} \geq p$$

Practical Implication: Prevents the algorithm from identifying turning points that are too close to each other.

Complete Cycles

Goal: Ensure the duration of a full cycle (peak-to-peak or trough-to-trough) respects the cycle duration (c).

Condition:

$$t_{\text{peak}2} - t_{\text{peak}1} \geq c \quad \text{and} \quad t_{\text{trough}2} - t_{\text{trough}1} \geq c$$

Effect: Filters out fluctuations that are too short to represent meaningful cycles.

Workflow of the BBQ Algorithm

Step 1: Detect Turning Points (Using ltpoint)

- Identify potential peaks and troughs using local maxima/minima rules.

Step 2: Filter by phase (stpoint)

- Ensure turning points are separated by at least p periods.

Step 3: Filter by cycle (otpoint)

- Ensure complete cycles are separated by at least c periods.

Step 4: Finalize results

- Retain valid peaks (+1) and troughs (-1).
- Assign 0 to other points.

Example: Application of BBQ

Consider a quarterly GDP series:

$$x_t = \{100, 105, 102, 98, 110, 107, 100, 95, 102, 110\}$$

Parameters:

- Window (w) = 2
- Phase (p) = 2
- Cycle (c) = 5

Output:

- Peaks: $t = 2, 9$
- Troughs: $t = 4, 8$
- Valid cycle: Peak-to-Peak = $t_9 - t_2 = 7 \geq c$

Strengths and limitations

Strengths:

- Objective and consistent methodology.
- Works directly on raw data.
- No need for explicit trend-cycle decomposition.

Limitations:

- Sensitive to noise in data.
- Tends to generate symmetric cycles.
- May require preprocessing (e.g., seasonal adjustment).

Cycles

At a minimum, three key aspects must be considered for any cycle:

1. How do I define it?
2. How do I date it?
3. **Filter the data.** If necessary, how do I filter the data to differentiate between the trend and short-term, medium-term, and long-term components?

Filter

The **Hodrick-Prescott (HP) Filter** is a widely used econometric tool for separating a time series into two components: trend component and cyclical component.

The filter is particularly popular in macroeconomics and finance for analyzing business cycles, economic growth trends, and asset price movements.

- Risk analysis: identify deviations from expected trends in economic indicators

Mathematical Formulation

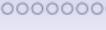
Objective Function

Given a time series y_t (e.g., GDP, stock prices), the HP filter solves the following minimization problem:

$$\min_{\tau_t} \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2$$

Where:

- τ_t : Trend component.
- $y_t - \tau_t$: Cyclical component.
- λ : Smoothing parameter controlling the trade-off between trend smoothness and data fit.



Mathematical Formulation (Interpretation)

- The first term $(y_t - \tau_t)^2$ ensures the trend closely follows the data.
- The second term penalizes rapid changes in the trend's slope, enforcing smoothness.

The Smoothing Parameter λ

- **Small λ :** Trend closely follows the data (less smoothing).
- **Large λ :** Trend becomes smoother, capturing fewer short-term fluctuations.

Common Values of λ :

- Quarterly data: $\lambda = 1600$
- Annual data: $\lambda = 100$
- Monthly data: $\lambda = 14400$

Limitations

Key Issues:

- **End-Point Bias:**
 - Poor performance at the beginning and end of the sample.
- **Arbitrary Choice of λ :**
 - Results can be sensitive to the choice of smoothing parameter.
- **Misinterpretation of Cycles:**
 - The cyclical component may include noise or artifacts.
Hamilton (2018) "Why You Should Never Use the Hodrick-Prescott Filter."
- **Deterministic Trend Assumption:**
 - Assumes trends evolve smoothly, which may not hold in all cases.

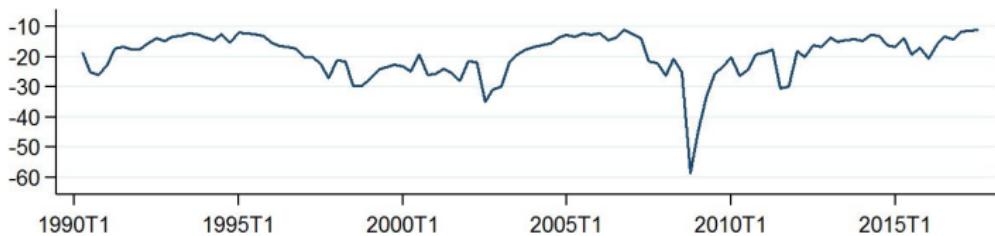
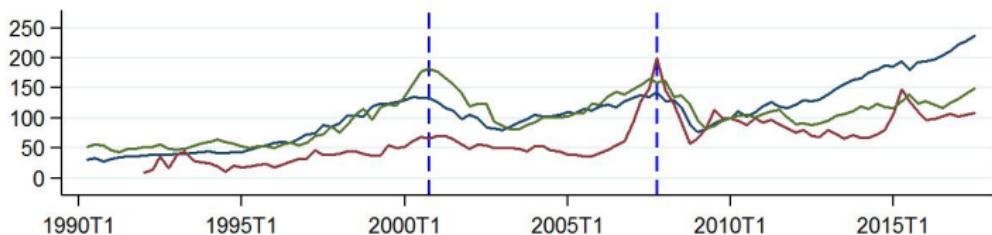
Cycles financiers?

Le VIX est une mesure en temps réel de la volatilité implicite du SP 500. Son calcul repose sur la volatilité des options d'échéance 30 jours sur le SP 500.

Cette volatilité implicite est elle même obtenue à partir de la médiane entre le prix d'achat et le prix de vente (bid/ask spread) de toutes les options, quel que soit leur prix d'exercice.

VIX

Les différents cycles financiers



Indices boursiers, 2010 T1 = 100. VIX: moyenne trimestrielle.

Financial Cycles

Original definition from Borio (BIS, 2012):

It denotes self-reinforcing interactions between perceptions of value and risk, attitudes towards risk and financing constraints, which translate into booms followed by busts. These interactions can amplify economic fluctuations and possibly lead to serious financial distress and economic dislocations.

National vs global financial cycle (Rey, 2013)

Financial Cycles: First Step

Variable selection: credit to the economy, credit divided by GDP, real estate prices, stock market prices

Variable transformation:

- Variables in levels should be deflated unless they are ratios (e.g., credit/GDP).
- Logarithmic transformation?
- Growth rates to normalize?

Financial Cycles: Second Step

Apply a statistical filter to extract the cyclical component from quarterly data.

Hodrick-Prescott Filter:

- $\lambda = 1600$ for short-to-medium-term cycles, suitable for economic cycles (Hardin, Pagan, 2000).
- $\lambda = 400,000$ for medium-to-long-term cycles, suitable for financial cycles (Adarov, 2018).
- For annual data, $\lambda = 6.5$ for short-to-medium-term cycles, 100 for medium-to-long-term cycles (Jorda et al., 2016).

Financial Cycles: Third Step

If working with variables in levels, it is important to normalize the cyclical component of the selected variables.

This allows them to be expressed on the same scale, facilitating aggregation into a single composite indicator.

Several approaches exist. For example, the BIS method (Drehmann et al., 2012): each series is divided (normalized) by its value at a specific date, the same for all series under study (e.g., 1985Q1).

Financial Cycles: Fourth Step

Aggregate the series to compute a composite financial cycle indicator.

1. Simple arithmetic mean of the normalized cyclical components of the financial cycle variables (BIS).
2. Weighted mean of the normalized cyclical components (weights derived using PCA on these variables - Bouvatier et al., 2015).

Financial Cycles: Fifth Step

Dating turning points in the financial cycle using the BBQ algorithm (Hardin, Pagan, 2002).

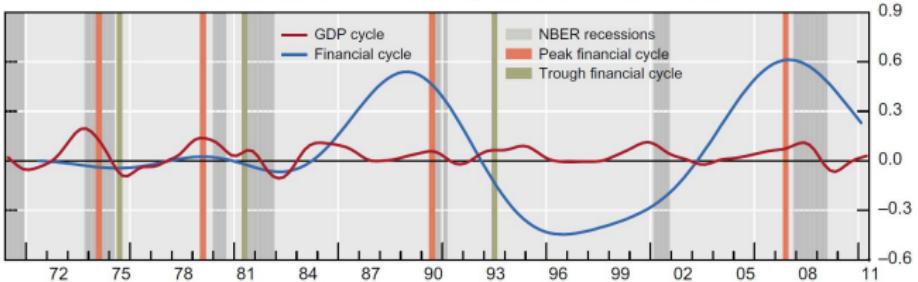
- Minimum cycle duration: 5 quarters for short-to-medium-term cycles, 20 quarters for medium-to-long-term cycles.
- Minimum phase duration: 2 quarters for the former, 8 quarters for the latter.
- Search window for local maxima (peaks) and minima (troughs): 5 quarters centered on t for the former, 18 quarters for the latter.

Financial Cycles

184

C. Borio / Journal of Banking & Finance 45 (2014) 182–198

The financial and business cycles in the United States



Graph 1. The financial and business cycles in the United States. Orange and green bars indicate peaks and troughs of the financial cycle measured by the component series (credit, the credit to GDP ratio and house prices) using the turning-point method. The blue line traces the financial cycle measured by the medium-term cycle in the component series using frequency-based filters. The red line traces the GDP cycle identified by the traditional shorter-term used to measure the business cycle. The amplitude of the cycles traced by the olive and red lines are not directly comparable. Source: Drehmann et al. (2014).

Financial Cycles

Some stylized facts. 5 from Borio (2014):

1. It is most parsimoniously described in terms of credit and property prices
2. It has a much lower frequency than the traditional cycle
3. Its peaks are closely associated with financial crises
4. It helps detect financial distress risks with a good lead in real time
5. Its length and amplitude depend on policy regimes

Links between business and financial cycles

Table 3

Business cycles with intense financial cycles.

Recessions associated with financial disruptions	Output					Financial variables		
	Number of events	Duration	Amplitude	Cumulative loss	Slope	Credit	House price	Equity price
A. Recessions without credit crunches	204	3.82	-2.74	-4.43	-0.84	1.13	-3.33	-12.98
Recessions with Credit Crunches	36	3.97	-3.62*	-7.80**	-0.83	-8.35***	-5.24*	-7.22
Recessions with Severe Credit Crunches	20	3.85	-3.54	-8.26	-0.96	-13.51***	-6.30*	-2.63
B. Recessions without house price busts	95	3.38	-1.96	-3.08	-0.60	1.25	-1.53	-14.42
Recessions with house price busts	46	4.74***	-2.76**	-7.29***	-0.62	-0.08***	-10.47***	-9.65
Recessions with severe house price busts	26	5.04**	-2.76**	-5.86***	-0.78	-2.28***	-10.74***	-9.65
C. Recessions without equity price busts	144	3.55	-2.15	-3.41	-0.57	0.74	-2.36	-4.55
Recessions with equity price busts	76	4.21**	-3.85***	-6.85***	-0.98***	1.06	-4.89**	-31.58***
Recessions with severe equity price busts	38	4.47**	-5.17***	-9.73***	-1.27***	1.10	-4.86*	-40.52***
Recoveries associated with financial booms	Output					Financial variables		
	Number of events	Duration	Amplitude	Slope		Credit	House price	Equity price
A. Recoveries without credit booms	216	5.22	3.94	1.07	1.38	-0.83	14.78	
Recoveries with credit booms	21	4.15	8.84***	2.13***	8.86***	3.26	14.48	
Recoveries with strong credit booms	11	4.45	10.41***	2.21***	13.02***	4.76	42.75**	
B. Recoveries without house price booms	126	4.76	2.89	0.75	0.81	-1.35	14.46	
Recoveries with house price booms	14	2.29***	6.14***	1.35***	5.99**	7.96***	15.23	
Recoveries with strong house price booms	9	2.44***	6.65***	1.59**	5.93	7.91***	18.37	
C. Recoveries without equity price booms	161	4.89	3.96	1.11	2.66	-0.38	5.36	
Recoveries with equity price booms	55	4.69	4.36	1.13	0.86	-0.44	38.53***	
Recoveries with strong equity price booms	30	5.18	4.46	1.21	0.57	-0.44	47.90***	

Notes: All statistics except "Duration" correspond to sample medians. For "Duration" means are shown. Booms are the top 25% of upturns calculated by the amplitude. Disruptions (Crunches, Busts, and Collapses) are the worst 25% of downturns calculated by the amplitude. Significance refers to the difference between recessions (recoveries) with and without financial disruptions (booms). For definitions of other statistics see notes to Tables 1 and 2.

* Implies significance at the 10% level.

** Implies significance at the 5% level.

Surprises monétaires: Illustration de la NLP

Cas des surprises monétaires avec Aruboa et Drechsel (WP, 2024).

Le titre est parlant: *Identifying Monetary Policy Shocks: A Natural Language Approach*

Motivation générale du papier de Aruboa et Drechsel (WP, 2024)

Objectif : Étudier comment les variations des taux d'intérêt Δi_t affectent les variables macroéconomiques (Y_t).

L'équation classique de Romer et Romer (2004) est donnée par :

$$\Delta i_t = \alpha + \beta i_{t-1} + \gamma X_t + \varepsilon_t^{RR}$$

où X_t contient les prévisions économiques issues des "Greenbooks" du FOMC.

Dans l'approche de Romer et Romer (2004), $R^2 = 0.5$. Les surprises monétaires sont dans le résidu ε_t^{RR} !

L'étude cherche à améliorer Romer et Romer (2004) en utilisant :

- **NLP (Natural Language Processing)** pour transformer les textes des documents du FOMC en données exploitables.
- **Machine learning** pour inclure de larges quantités de données textuelles dans les régressions, de manière linéaire ou non.

Principaux résultats

1. Meilleure explication des décisions de la FED

$$R^2 = 0.94 \text{ (approche actuelle)}$$

contre

$$R^2 = 0.5 \text{ (approche Romer et Romer)}$$

2. Les chocs de politique monétaire sont définis comme des "surprises" pour les analystes du FOMC, basées sur des préoccupations à long terme.

Que sont les chocs de politique monétaire ?

Les chocs peuvent être interprétés comme des "surprises pour le staff".

Exemple : Novembre 1994, le plus grand choc de resserrement.

- Le staff de la Fed anticipait une hausse de 50 points de base.
- Alan Greenspan a plaidé pour une hausse de 75 points de base.
- Choc estimé : 21 points de base de resserrement.

Étape 1 : Traitement du texte brut

Processus :

- Téléchargement des documents des réunions du FOMC (Beigebook, Tealbook A, etc.).
- Extraction des chocs directement à partir des réunions, en commençant en 1982.

Étape 2 : Identification des concepts économiques

Tous les termes économiques discutés sont stockés en simples, doubles et triples occurrences. Les concepts les plus fréquemment discutés sont ensuite sélectionnés :



Étape 3 : Construction des scores de sentiment

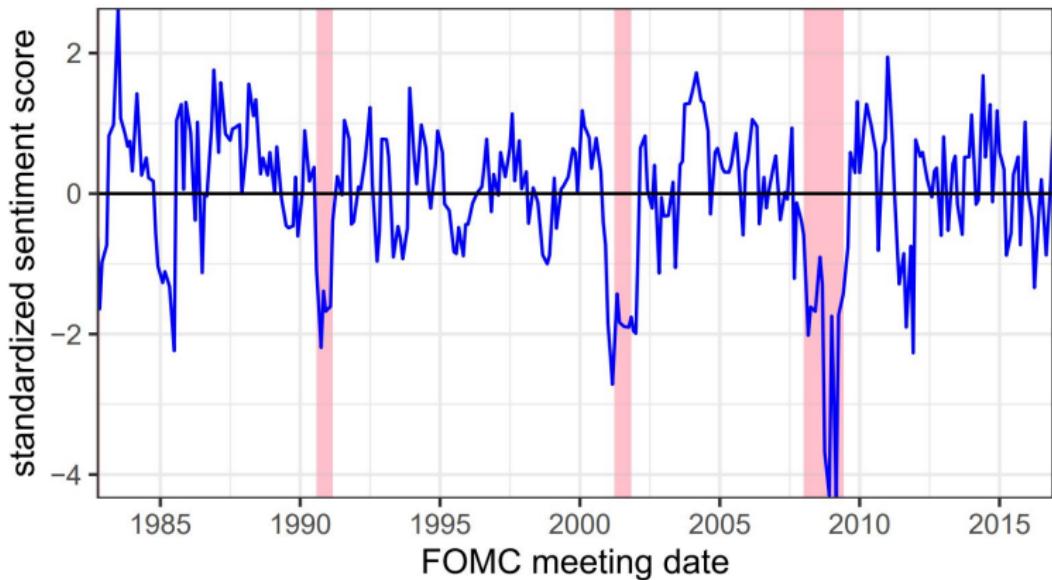
Inspiré par Hassan et al. (2022), chaque terme économique est associé à un score de sentiment en fonction des mots précédents et suivants. +1 pour chaque terme positif, -1 pour chaque terme négatif.

$$\text{Score total} = \frac{\sum(\text{mots positifs} - \text{mots négatifs})}{\text{Nombre total de mots}} \quad (40)$$

Suite de l'étape 3

Positive	Negative
adequate	adversely
advantage	aggravate
benefit	bad
boost	burdensome
confident	collapse
conducive	concerning
desirable	decline
diligent	deficient
encouraging	eroded
excellent	exacerbate
...	...

Suite de l'étape 3



Pour l'analyse de sentiments sur le mot "economic activity"

Étape 4 : Régression

L'équation de la régression est donnée par :

$$\Delta i_t = \alpha + \beta i_{t-1} + \Gamma(X_t, Z_t) + \varepsilon_t^*$$

où :

- X_t représente les prévisions numériques (toutes variables et lags).
- Z_t représente les indicateurs de sentiment avec lags.
- $\Gamma(\cdot)$ capture les non-linéarités.

Problème majeur: il y a 3226 variables pour 210 observations.

LASSO ou RIDGE.

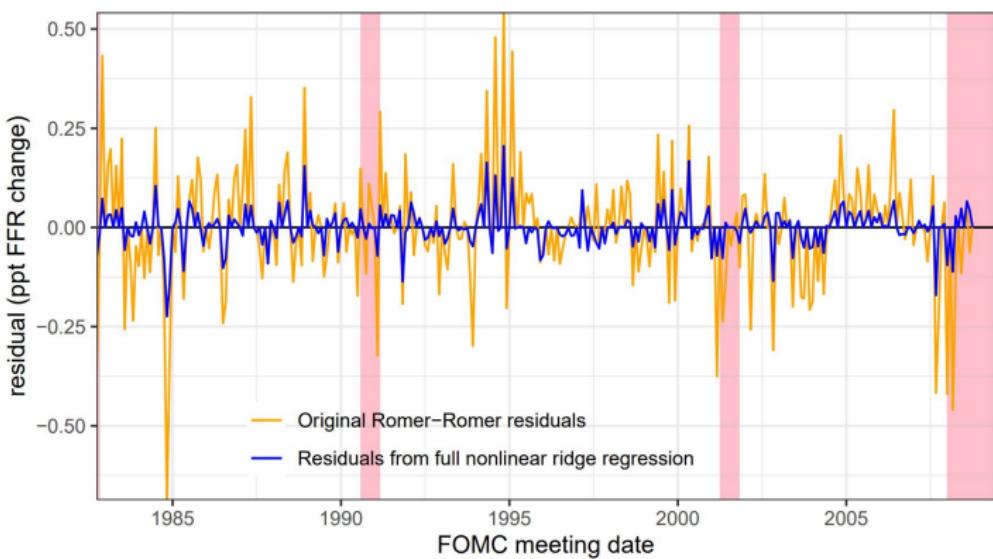
Résultats de la procédure d'identification

	Left hand side: Greenbook unemployment rate forecast errors							
	(1) current quarter	(2) 1-quarter ahead	(3) 1-year ahead	(4) 2-years ahead	(5) current quarter	(6) 1-quarter ahead	(7) 1-year ahead	(8) 2-years ahead
First PC of all sentiments	-0.029* [0.016]	-0.114** [0.049]	-0.445** [0.190]	-0.622** [0.238]				
Economic activity sentiment					-0.026 [0.016]	-0.098** [0.048]	-0.285* [0.165]	-0.363** [0.171]
Constant	-0.019 [0.014]	-0.070** [0.033]	-0.082 [0.121]	0.059 [0.201]	-0.019 [0.014]	-0.069** [0.035]	-0.077 [0.145]	0.160 [0.258]
R-squared	0.045	0.149	0.248	0.208	0.033	0.097	0.090	0.055
Obs	210	210	210	62	210	210	210	62

Résultats de la procédure d'identification

	Number of regressors	(1) R^2 with 10-word sentiment (main specification)	(2) R^2 with 5-word sentiment (robustness)
Romer-Romer original OLS with subset of forecasts	19	0.50	
Ridge with extended set of forecasts	133	0.55	
Ridge with all forecasts & sentiments (linear)	429	0.65	0.66
Ridge with all forecasts & sentiments (nonlinear)	858	0.75	0.77
Ridge with all forecasts & sentiments (linear with lags)	1,613	0.87	0.88
Ridge with all forecasts & sentiments (nonlinear with lags)	3,226	0.94	0.95

Chocs de politique monétaire estimés



Les chocs de politique monétaire estimés par notre méthode sont fortement corrélés avec les surprises de marché. 83% de corrélation entre les 2 méthodes.

Samuel Ligoni  re

(Ann  e universitaire 2024-2025)

M2 Advanced Econometrics

185 / 185