

Calibration

Matthieu Garcin

ESILV

2023-2024



Note

- 50% mini-projet en groupes de deux étudiants (sujet TD 5 à 7) : le rendu est le code et un rapport en **LaTeX**
 - 12.5% participation/implication (dont présentation d'articles pendant les TD 1 à 6)
 - 30% examen écrit sur le cours
 - 7.5% précision de la réponse à un *jeu de trading d'option* (sujet TD 3 et 4)
-
- 15h de CM, 21h de TD (TD en séances de 3h).
 - TD : apportez votre ordinateur.
 - Code : utilisation de librairies prohibée.
 - Questions : prioritairement en cours/TD, sinon sur rendez-vous.
 - Stages : liste fonds sur DVO.

Modalités - Présentations d'articles

A l'oral en début de chaque TD (calendrier à respecter), 5 à 10 minutes, travail individuel.

Bonus sur la note de participation/implication, avec un coefficient dépendant de ce qui est présenté (les éléments sont additifs) :

- présentation de l'article : coef 1 ;
- reproduction des résultats sur d'autres données : coef 1 ;
- proposition et implémentation d'améliorations : coef 2.

Certaines présentations seront des revues de littératures et suivent une notation indépendante : coef 2.

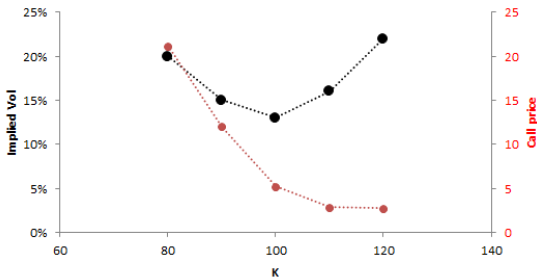
Problème pratique :

- On observe le prix de 5 options d'achat sur le même sous-jacent : $S_0 = 100$, $r = 0$, $T = 1$, $K \in \{80, 90, 100, 110, 120\}$. Leur prix est respectivement 21.19, 12.02, 5.18, 2.85, 2.74.
- En tant qu'investisseur, on nous propose d'acheter au prix de 7.00 une autre option d'achat sur le même sous-jacent, avec $K = 102$ et $T = 2$.
- Opportunité à saisir ?
- On veut comparer prix proposé à prix fourni par modèle de Black-Scholes, mais avec quelle volatilité ?

Introduction

Plusieurs approches :

- Naturellement, on regarde l'option de maturité 1 an de strike le plus proche ($K = 100$) et on estime la volatilité selon modèle de Black-Scholes, ce qui fournit une volatilité de 13%. On applique cette même volatilité à l'option 2 ans de strike 102, ce qui fournit, toujours par Black-Scholes, 6.44 : **le prix de 7 est trop élevé.**
- Possibilité d'estimer un paramètre de vol $\hat{\sigma}(K)$ pour chaque strike K . Mais faut-il prendre :
 - $\frac{4\hat{\sigma}(100) + \hat{\sigma}(110)}{5} = 13.6\%$, soit un prix de 6.78 : **7 est trop élevé.**
 - $\frac{1}{5} \sum_{i=-2}^2 \hat{\sigma}(100 + i \times 10) = 17.2\%$, soit un prix de 8.81 : **7 est bon marché.**



Introduction

Quelques questions liées à cet exemple :

- Pour des options ou des modèles plus sophistiqués, comment estimer les paramètres ?
 - Quel problème d'optimisation à résoudre ?
 - Quel algorithme utiliser pour le résoudre ?
- Comment interpoler/extrapoler le *smile* de volatilité implicite ?
- Si le modèle de Black-Scholes nécessite de tels écarts de paramètres selon le strike, la dynamique brownienne est-elle bien spécifiée ?
- Un risque de modèle peut surgir de cette mauvaise spécification, mais aussi :
 - de problèmes inverses mal posés (si la réciproque du prix en fonction du paramètre est discontinue) qui nécessitent d'être régularisés, c'est-à-dire non pas d'être transformés en problèmes bien posés mais de tirer le maximum d'information avec des outils adaptés ;
 - des données (univers d'options) trop peu nombreuses ou non-cohérentes avec le besoin (proxy de sous-jacent) ;
 - d'algorithmes d'optimisation peu stables, piégés par des minima locaux ;
 - de méthodes de calcul de prix trop incertaines (intervalles de confiance d'un Monte-Carlo, approximations numériques d'un schéma aux différences finies).

- Une autre source de risque de modèle est la présence de bruit qui introduit de l'incertitude dans l'estimation. La différence entre estimation et calibration (entre financiers statisticiens et financiers probabilistes) vient de ce que la calibration suppose que le prix observé est le vrai prix, sans bruit.
- Mais les statistiques constituent l'outil pour trouver les paramètres d'un modèle probabiliste. Pourquoi s'en passer ? N'y a-t-il pas de lien entre les deux approches, entre la régularisation d'un problème mal posé et la régularisation en statistiques non-paramétriques ?

Objectifs du cours :

- savoir poser le problème d'optimisation (i.e. de calibration des paramètres) permettant de tirer le plus grand profit de l'information disponible ;
- maîtriser les techniques d'optimisation adaptées au problème (linéaire ou non, convexe ou non, avec des minima locaux ou non, de grande dimension ou non, avec des contraintes ou non...) ;
- maîtriser tous les aspects de la calibration de la volatilité implicite, dont la connaissance de modèles plus avancés que Black-Scholes ;
- comprendre l'articulation entre calibration et statistiques ;
- appliquer toutes ces notions à des modèles variés : modèles de valorisation ou modèles de prévision.

- 1 Algorithmes d'optimisation
 - Algorithmes déterministes
 - Algorithmes stochastiques
- 2 Calibration de volatilité implicite
 - Modèle de Black-Scholes, volatilité implicite, volatilité locale
 - Modèles à volatilité stochastique
 - Densité risque-neutre
 - Paramètres implicites d'un mouvement Lévy-stable fractionnaire
- 3 Théorie de la calibration
 - Régularisation
 - Estimation ou calibration ?
- 4 L'approche statistique
 - Régularisation et statistiques non-paramétriques
- 5 Exemples de calibrations
 - Courbe de taux

Écosystème des algorithmes d'optimisation

On s'intéresse uniquement ici à l'optimisation numérique (paramètres dans \mathbb{R}). Existent aussi l'optimisation combinatoire ou discrète (paramètres dans un ensemble fini ou dénombrable de valeurs), la commande (ou contrôle) optimale (le "paramètre" est une fonction), l'optimisation stochastique (données aléatoires, à ne pas confondre avec algorithmes stochastiques)...

Deux cas particuliers éloignés du genre de modèles que l'on utilise en finance :

- Problème et contraintes linéaires (i.e. programmation linéaire) : résolution avec méthode du simplexe ou algorithme des points intérieurs
- Problème quadratique et contraintes linéaires (i.e. programmation quadratique ; exemple : problème de sélection de portefeuille) : si contraintes d'égalité, revient à problème linéaire, sinon problème beaucoup plus difficile, résolution par des extensions du simplexe ou des points intérieurs.

Écosystème des algorithmes d'optimisation

Des cas plus généraux, avec un problème non-linéaire mais sans contrainte ; méthodes heuristiques, c'est-à-dire non exactes mais rapides, adaptées à un problème particulier (sauf les méthodes métaheuristiques, plus générales, que nous voyons ici) :

- en utilisant la dérivée (problème inverse mal posé...) : gradient, **Newton-Raphson**,¹ Gauss-Newton, Levenberg-Marquardt...
- sans utiliser la dérivée : déterministes (**Nelder-Mead, dichotomie**), stochastiques² (**recuit simulé, recherche tabou, algorithmes génétiques/évolutionnaires, essais particuliers**).

Pour finir, des problèmes non-linéaires mais avec contraintes : gradient projeté, algorithme d'Uzawa...

1. Algorithmes en gras présentés dans cette partie.

2. Attention, s'il est mal paramétré, un algorithme stochastique lancé deux fois peut fournir deux résultats différents.

1 Algorithmes d'optimisation

- Algorithmes déterministes
 - Méthode de Newton-Raphson
 - Dichotomie
 - Algorithme de Nelder-Mead
- Algorithmes stochastiques

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Méthode de Newton-Raphson

- On cherche le minimum de h , fonction $\mathcal{C}^2(\mathbb{R})$. Cela revient à chercher le zéro de h' .
- Abordons le problème progressivement : Soit une fonction continûment dérivable définie dans un espace de dimension 1 : $f : \mathbb{R} \rightarrow \mathbb{R}$.

$$f'(x_i) = \frac{f(x_i) - 0}{x_i - x_{i+1}},$$

donc :

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}.$$

En pratique

Si l'expression exacte de h' et $f' = h''$ est trop compliquée, on peut éventuellement remplacer par une dérivée empirique (différences finies), en prenant un très petit pas. On peut aussi remplacer $f'(x_i)$ par $\frac{f(x_i) - f(x_{i-1}))}{x_i - x_{i-1}}$, pour une moins bonne convergence : c'est la **méthode de la sécante**.

Méthode de Newton-Raphson

Exemple avec 4 paramètres à optimiser :

- On cherche minimum d'une fonction non-linéaire mais deux fois continûment dérivable : $(a, b, c, d) \mapsto g_t(a, b, c, d)$.
- Objectif : minimiser l'écart quadratique entre observation $p(t)$ (par exemple un prix) et modèle, entre les dates T_0 et T_1 :

$$(a, b, c, d) = \min_{(a', b', c', d')} \sum_{t=T_0}^{T_1} [g_t(a', b', c', d') - p_t]^2.$$

- Recherche de racine : paramètres (a', b', c', d') annulant la fonction F suivante :

$$F(a', b', c', d') = \sum_{t=T_0}^{T_1} [g_t(a', b', c', d') - p_t] \begin{pmatrix} \partial_a g_t \\ \partial_b g_t \\ \partial_c g_t \\ \partial_d g_t \end{pmatrix} (a', b', c', d').$$

Ouf!

Pas besoin d'avoir $t \mapsto p_t$ dérivable, seule la dérivabilité de g en fonction de chacun de ses paramètres (mais pas de t) est requise : on a intégré selon t et dérivé selon a, b, c et d .

Méthode de Newton-Raphson

Posons $F(a', b', c', d') = \begin{pmatrix} F^1 \\ F^2 \\ F^3 \\ F^4 \end{pmatrix} (a', b', c', d')$.

On définit ensuite la Jacobienne, J , de F ainsi :

$$J = \begin{pmatrix} D^1 F^1 & D^2 F^1 & D^3 F^1 & D^4 F^1 \\ D^1 F^2 & D^2 F^2 & D^3 F^2 & D^4 F^2 \\ D^1 F^3 & D^2 F^3 & D^3 F^3 & D^4 F^3 \\ D^1 F^4 & D^2 F^4 & D^3 F^4 & D^4 F^4 \end{pmatrix},$$

où $D^j F^i$ est la différentielle de F^i selon la j -ième variable.

Aux points où cette matrice est inversible, on définit la suite des paramètres $(a^i, b^i, c^i, d^i)_{i \in \mathbb{N}}$ – notée $(X_i)_{i \in \mathbb{N}}$, qui converge vers le zéro de F , par la relation de récurrence :

$$X_{i+1} = X_i - J^{-1}(X_i)F(X_i).$$

Méthode de Newton-Raphson

Synthèse de l'exemple dans \mathbb{R}^4

Comme $D^j F^i = \sum_{t=T_0}^{T_1} \partial_j g_t \partial_i g_t + (g_t - p_t) \partial_{i,j} g_t$, on a :

$$J = \sum_{t=T_0}^{T_1} \nabla g_t \times \nabla g_t^T + (g_t - p_t) \cdot H(g_t),$$

où

- ∇g_t est le vecteur gradient de g_t ,
- T l'opérateur de transposition,
- \times la loi de multiplication matricielle,
- \cdot la multiplication par un scalaire,
- $H(g_t)$ la matrice hessienne de g_t , symétrique selon le théorème de Schwarz (car $(a, b, c, d) \mapsto g_t(a, b, c, d)$ deux fois dérivable).

Méthode de Newton-Raphson

Attention

Inversion de J ou de f' : attention aux instabilités. Lorsque $f'(x_i)$ est proche de 0, x_{i+1} s'éloigne de x_i .

Conséquences :

- Bien choisir le point initial x_0 , pas trop loin de l'optimum.
- Si f' décroît plus vite que f près de l'optimum, cette méthode n'est pas la bonne (coupler éventuellement avec une autre, type dichotomie) : boucle infinie (sauf si on a limité le nombre d'itérations en plus du critère d'arrêt de précision).
- Si x_i sort du domaine de définition de f (e.g. $f = \ln$ avec $x_0 \geq e$) ou que $f'(x_i) = 0$ (e.g. $f : x \mapsto x^3 - 1$ et $x_0 = -2^{-1/3} \approx -0.7937$)³, on ne peut même plus itérer.
- Cas de divergence grossière : $f : x \mapsto 1 - \exp(-x^2)$, avec $x_0 \in]-\infty, -\xi] \cup [\xi, +\infty[$, où $\xi \approx 1.529$, solution de $\exp(x^2) - 4x^2 - 1 = 0$.
- Quelques cas pathologiques : $x \mapsto |x|^{1/2}$, certes non dérivable en 0, $x_{i+1} = -x_i$ pour tout i , tant que $x_0 \neq 0$.

3. En prenant $x_0 = -0.7937$, converge avec erreur $< 10^{-4}$ en 70 itérations, réduit à 24 itérations si on prend $x_0 = -0.8$.

Méthode de Newton-Raphson – convergence

Soit z un zéro de f . Alors, par la formule de Taylor-Lagrange :

$$0 = f(z) = f(x_i) + f'(x_i)(z - x_i) + \frac{1}{2}f''(\xi_i)(z - x_i)^2,$$

où $\xi_i \in [z, x_i]$. En itérant, on obtient un écart entre racine et estimé qui est :

$$x_{i+1} - z = x_i - \frac{f(x_i)}{f'(x_i)} - z = \frac{f''(\xi_i)}{2f'(x_i)}(x_i - z)^2.$$

Avec $f'(x)$ et $f''(x)$ bornées pour x dans un intervalle, on a :

$$|x_{i+1} - z| \leq K|x_i - z|^2$$

puis, par récurrence :

$$(K|x_i - z|) \leq (K|x_0 - z|)^{2^i}.$$

La convergence est quadratique ($|x_{i+1} - z|/|x_i - z|^2 \leq K$), si tant est que $K|x_0 - z| < 1$ (condition suffisante). Par exemple, on peut prendre

$$K = \max_{x \in [z, x_0]} |f''(x)|/2 \min_{x \in [z, x_0]} |f'(x)|.$$

Méthode de Newton-Raphson – convergence

Importance du choix de x_0 pour minimiser K . On peut alors estimer nombre d'itérations suffisantes n_ε pour avoir convergence $|x_i - z| \leq \varepsilon : n_\varepsilon = \log_2 \left(\frac{\ln(\varepsilon K)}{\ln(K|x_0 - z|)} \right)$.

Exemples :

- Si $f(x) = x$, $K = 0$, convergence immédiate.
- Si $f(x) = x^2$, K est infini car la dérivée s'annule à la racine de f : convergence rapide au début puis qui ralentit très fortement à proximité de la racine et peut même diverger (coupler avec autre algorithme pour ne pas trop s'éloigner). Il faut forte pente et faible convexité/concavité pour converger vite.
- Si $f(x) = x^2 - 1$, avec $x_0 > 1$, K vaut $1/2$ donc convergence assurée par ce qui précède (en pratique meilleur, car le K ne reste pas le même à chaque itération et exagère le reste du développement de Taylor) pour tout x_0 tel que $x_0 - 1 < 2$.
- Même exemple avec $x_0 < 1$. Si $x_0 > 0$, convergence vers 1 à rythme d'autant plus faible que x_0 est proche de 0. Si $x_0 < 0$, convergence vers -1. Si $x_0 = 0$, K infini et divergence grossière.

Méthode de Newton-Raphson

Dans le cas de l'estimation de la volatilité implicite pour N options sur le même sous-jacent et de même échéance, on cherche $(\sigma(K_i))_{1 \leq i \leq N}$ tel que $\forall i, BS(\sigma(K_i), K_i) = C(K_i)$, où $BS(\sigma(K_i))$ est le prix Black-Scholes de paramètre de vol $\sigma(K_i)$ et de strike K_i et où $C(K_i)$ est le prix du call tel qu'observé sur le marché.

- On cherche alors le zéro de $(\sigma(K_1), \dots, \sigma(K_N)) \in \mathbb{R}^N \mapsto \sum_{i=1}^N |BS(\sigma(K_i), K_i) - C(K_i)|$ ou bien on peut diviser ce problème en N problèmes de dimension 1.
- Si on observait plus de prix qu'on ne recherchait de $\sigma(K_i)$: plus d'équations que d'inconnues, potentiellement aucune solution sauf en utilisant approche statistique.
- Si on recherchait plus de $\sigma(K_i)$ qu'on observait de prix : plus d'inconnues que d'équations, problème mal posé, admet une infinité de solutions ; choisir la meilleur selon un certain critère de régularité (tomographie).

1 Algorithmes d'optimisation

- Algorithmes déterministes

 - Méthode de Newton-Raphson

 - Dichotomie

 - Algorithme de Nelder-Mead

- Algorithmes stochastiques

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Dichotomie

Appelée aussi méthode de la **bissection**. Principe :

- Initialisation : $a_0 = a$ et $b_0 = b$ tel que $f(a) \times f(b) < 0$.
- Itération : on a $f(a_i) \times f(b_i) < 0$. Soit $c_i = (a_i + b_i)/2$. Alors
 - Si $f(c_i) = 0$: fin de l'algorithme, la solution est trouvée.
 - Si $f(c_i) \times f(a_i) < 0$: $a_{i+1} = a_i$ et $b_{i+1} = c_i$.
 - Si $f(c_i) \times f(a_i) > 0$: $a_{i+1} = c_i$ et $b_{i+1} = b_i$.
- Converge si fonction est continue (théorème des valeurs intermédiaires). A l'itération i , on sait simplement que racine dans (a_i, b_i) . Rétrécissement de cet intervalle permet estimation.
- Convergence linéaire car $(b_{i+1} - a_{i+1})/(b_i - a_i) = 1/2$.
- Pour condition d'arrêt $b_i - a_i \leq \varepsilon$, on sait d'avance combien d'itérations sont suffisantes (et nécessaires si on ne tombe pas sur racine par chance) car $b_i - a_i = (b - a)2^{-i}$: il suffit d'avoir $i \geq \frac{1}{\ln(2)} \ln\left(\frac{b-a}{\varepsilon}\right)$.

En pratique :

- Si dérivée nulle à la racine, là encore difficile de trouver racine, car potentiellement pas de points pour lesquels fonction change de signe ($f(x) = x^2$).
- Vitesse de convergence moins bonne que Newton-Raphson, mais pas de divergence grossière ni de contraintes sur la dérivabilité.

1 Algorithmes d'optimisation

- Algorithmes déterministes
 - Méthode de Newton-Raphson
 - Dichotomie
 - Algorithme de Nelder-Mead
- Algorithmes stochastiques

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Nelder-Mead – dimension 1

Dichotomie nécessite la connaissance de a_0 et b_0 tel que $f(a_0) \times f(b_0) < 0$. L'algorithme de Nelder-Mead (en dimension 1) permet de contourner ce problème (on cherche le minimum de f ; dans le cas de recherche de racine de g , on a $f = |g'|$) :

- Initialisation : On choisit x_0^1 et x_0^2 quelconques et on ré-indexe de sorte que l'on ait $f(x_0^1) \leq f(x_0^2)$.
- Itération : $f(x_i^1) \leq f(x_i^2)$. On détermine $x_i^r = x_i^1 + (x_i^1 - x_i^2)$ la réflexion de x_i^2 par rapport à x_i^1 . Autrement dit, on sort de notre intervalle.
 - Si $f(x_i^r) < f(x_i^1)$, on va vouloir étirer notre intervalle pour que x_i^r tombe en son milieu : $x_i^e = x_i^1 + 2(x_i^1 - x_i^2)$. On vérifie que cela est pertinent :
 - si $f(x_i^e) \leq f(x_i^r)$ on étire effectivement $x_{i+1}^1 = x_i^e$ et $x_{i+1}^2 = x_i^1$;
 - sinon, on se contente de la réflexion $x_{i+1}^1 = x_i^r$ et $x_{i+1}^2 = x_i^1$.
 - Si $f(x_i^r) > f(x_i^1)$, on contracte l'intervalle avec $x_{i+1}^1 = x_i^1$ et $x_{i+1}^2 = x_i^2 + \frac{1}{2}(x_i^1 - x_i^2)$ (à ré-indexer).

La dernière étape (contraction) est digne d'une dichotomie et permet d'espérer converger linéairement si fonction pas trop compliquée et optimum pas trop loin de l'intervalle initial. L'étirement de l'intervalle permet au contraire de déplacer cet intervalle avec augmentation potentielle de sa taille (conditions font que si $x_{i+1}^1 = x_i^r$, on ne peut pas avoir boucle infinie avec $x_{i+2}^1 = x_i^2$, car $f(x_i^r) < f(x_i^1)$).

Nelder-Mead – simplexe

Vrai nom : simplexe de Nelder-Mead (rien à voir avec l'algorithme du simplexe en optimisation linéaire, où les côtés du simplexe correspondent à la saturation des contraintes).

Définition

Un **simplexe** de l'espace vectoriel \mathbb{R}^d est un polytope de $d + 1$ sommets. C'est-à-dire, si x_1, \dots, x_{d+1} sont les sommets, le simplexe est leur enveloppe convexe :

$$\left\{ x \in \mathbb{R}^d \left| x = \sum_{i=1}^{d+1} \lambda_i x_i, \sum_{i=1}^{d+1} \lambda_i = 1, \forall i \lambda_i \geq 0 \right. \right\}.$$

C'est une généralisation en dimension d du **triangle** (d'où le nom de simplexe : objet géométrique des plus simples).

- Définit un sous-ensemble de dimension d en utilisant le moins de sommets possibles.
- Chaque sommet a d arrêtes partant vers chacun des d autres sommets. Tous les sommets sont donc liés deux à deux.
- En dimension 1, le simplexe est un segment, en dimension 2 un triangle, en dimension 3 un tétraèdre, en dimension 4 un pentachore, et au-dessus, en dimension d , un d -simplexe.

Nelder-Mead – dimension ≥ 1

On cherche le minimum de f , fonction de \mathbb{R}^d .

- ➊ **Tri** : On a $d + 1$ vecteurs de paramètres x_1, \dots, x_{d+1} , que l'on trie et ré-indexe de sorte que $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{d+1})$.
- ➋ **Réflexion** : On calcule x_r , réflexion de x_{d+1} par rapport à x_0 , le barycentre de l'ensemble des vecteurs privé de x_{d+1} : $x_r = x_0 + (x_0 - x_{d+1})$.
- ➌ **Étirement** : Si $f(x_r) < f(x_1)$, on calcule le point étiré $x_e = x_0 + 2(x_0 - x_{d+1})$. S'il est meilleur que le point réfléchi, i.e. si $f(x_e) < f(x_r)$, on conserve le simplexe étiré en remplaçant x_{d+1} par x_e . Sinon, on remplace x_{d+1} par x_r (qui est meilleur que x_1).
- ➍ **Conservation de la réflexion** : Sinon, si $f(x_r) < f(x_d)$, on conserve la réflexion en remplaçant x_{d+1} par x_r .
- ➎ **Contraction** : Sinon, si $f(x_r) \geq f(x_d)$, on contracte le simplexe :
 - Contraction à l'intérieur : Si $f(x_d) \leq f(x_r) < f(x_{d+1})$, on calcule $x_c = \frac{1}{2}(x_0 + x_r)$. Si $f(x_c) \leq f(x_r)$, on remplace x_{d+1} par x_c .
 - Contraction à l'extérieur : Si $f(x_r) \geq f(x_{d+1})$, on calcule $x_c = \frac{1}{2}(x_0 + x_{d+1})$. Si $f(x_c) \leq f(x_{d+1})$, on remplace x_{d+1} par x_c .
 - Contraction vers x_1 : Dans les autres cas (x_c pas bon candidat), on contracte tous les vecteurs vers le meilleur : $\forall i \in \{2, \dots, d + 1\}, x_i = \frac{1}{2}(x_1 + x_i)$.

Chaque traitement de cas est justifié par le fait que le nouveau sommet est meilleur que l'un de ceux qui restent dans le simplexe, de sorte à ne pas remettre en jeu ce sommet à l'étape suivante (boucle infinie).

Nelder-Mead

- Algorithme se laissant piéger par minima locaux (plus qu'algorithmes aléatoires).
Une solution : lancer plusieurs fois l'algorithme avec des valeurs initiales différentes. On peut même lancer $d + 1$ fois l'algorithme et utiliser les $d + 1$ solutions pour initialiser un calcul supplémentaire.
- Difficulté du choix (arbitraire) des $d + 1$ vecteurs initiaux :
 - doit être suffisamment grand pour ne pas réduire la recherche à une recherche locale, même si le simplexe peut se déplacer (il se déplacera plutôt en suivant des courbes de niveau, s'il est de petite taille, et n'atteindra ainsi pas certains creux pour lesquels il faut d'abord monter).
 - dans le papier originel : choix d'un point x_1 puis autres vecteurs obtenus par translation à partir de x_1 dans toutes les directions d'un repère cartésien d'un pas fixe.
- Critère d'arrêt : quand le simplexe est de taille réduite (voir ci-après, le volume d'un simplexe), sous un seuil prédéterminé (x_1 , après tri, est alors la solution).
- Moins pertinent pour estimer un seul paramètre (la vol implicite d'une option) que plusieurs (vol implicite et autre paramètre de la dynamique?).
- Cet algorithme vise à minimiser une fonction, comme $(\sigma(K_1), \dots, \sigma(K_N)) \in \mathbb{R}^N \mapsto \sum_{i=1}^N |BS(\sigma(K_i), K_i) - C(K_i)|$ dans le cadre d'une calibration.

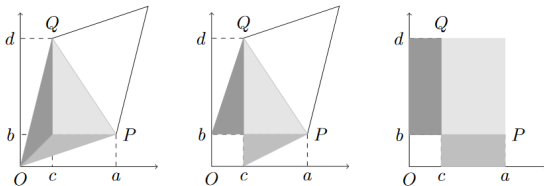
Nelder-Mead – critère d'arrêt (1/4)

Volume d -dimensionnel du simplexe :

- Une matrice $d \times d$, M , peut s'écrire en termes de colonnes : $M = (M_1, M_2, \dots, M_d)$.
- Le d -hypercube unitaire (défini par d vecteurs de l'espace orthonormé \mathbb{R}^d égaux aux vecteurs de base) est représenté par les d vecteurs utiles à sa construction et mis dans une matrice : I_d , matrice identité car vecteurs unitaires.
- $M \times I_d$ décrit les vecteurs de construction d'un d -paralléloèdre (généralisation du parallélépipède) $P = \left\{ x \in \mathbb{R}^d \mid x = \sum_{i=1}^d \lambda_i M_i, \forall i \lambda_i \in [0, 1] \right\}$: M transforme l'hypercube en paralléloèdre.
- Le déterminant de M , $\det(M)$, fournit le volume d -dimensionnel signé du paralléloèdre P (le signe indique si la transformation M conserve ou renverse l'orientation du cube initial).⁴
- Exemple : déterminant nul, vecteurs de construction du paralléloèdre liés, plusieurs manières de construire ce paralléloèdre, objet pas de dimension d , volume nul, transformation M non inversible.

Nelder-Mead – critère d'arrêt (2/4)

Déterminant est volume d -dimensionnel (aire si $d = 2$).



Source : Emmanuel Hallouin.

$$M = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Fig 1 : L'aire grise est la moitié de l'aire du parallélogramme.

Fig 2 : L'aire d'un triangle ne change pas si on bouge un sommet parallèlement à côté opposé.

Fig 3 : L'aire du parallélogramme est $ad - bc = |\det(M)|$.

Nelder-Mead – critère d'arrêt (3/4)

Proposition

Le volume d -dimensionnel du simplexe défini par les vecteurs x_1, \dots, x_{d+1} est $\left| \frac{1}{d!} \det(x_2 - x_1, \dots, x_d - x_1, x_{d+1} - x_1) \right|$.

- Le fait d'ôter x_1 à chaque vecteur permet de recentrer la figure (en comparaison, un parallélogramme est défini par deux vecteurs mais aussi par la donnée d'un sommet, O ci-après).
- A partir des d vecteurs $v_i = x_{i+1} - x_1$ (où $i \in \{1, \dots, d\}$), on peut définir des simplexes définis par les vecteurs w_0, \dots, w_d , avec $w_0 = x_1$ et $w_i = w_{i-1} + v_{\sigma(i)}$ pour σ , une permutation de $\{1, \dots, d\}$ donnée. Ces simplexes sont de même volume et ne se chevauchent pas. Il sont au nombre de $d!$, comme le nombre de permutations de $\{1, \dots, d\}$. Leur union est un polytope défini par les d vecteurs v_1, \dots, v_d . Ceci prouve la proposition.

Nelder-Mead – critère d'arrêt (4/4)

Outre le volume d -dimensionnel du simplexe :

- **Diamètre du simplexe** : L'espace vectoriel \mathbb{R}^d est muni de sa norme euclidienne. Le diamètre du simplexe défini par les vecteurs x_1, \dots, x_{d+1} est :

$$D(x_1, \dots, x_{d+1}) = \sup_{i \neq j} \|x_j - x_i\|.$$

- **Volume d -dimensionnel de l'hypersphère circonscrite** : L'hypersphère est une généralisation de la sphère en dimension 3 et du cercle en dimension 2. Pour un centre défini par un vecteur C et un rayon $r > 0$: $\{x \in \mathbb{R}^d \mid \|x - C\| \leq r\}$. Son volume $V_d(r) = \int_{-r}^r V_{d-1}(\sqrt{r^2 - x^2}) dx$ (théorème de Fubini) vaut :

$$V_d(r) = \frac{\pi^{d/2} r^d}{\Gamma\left(\frac{d}{2} + 1\right)},$$

avec $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ et $\Gamma(x+1) = x\Gamma(x)$. Ici, on peut donc prendre le critère $V_d\left(\frac{1}{2}D(x_1, \dots, x_{d+1})\right)$ inférieur à un certain seuil.

1 Algorithmes d'optimisation

- Algorithmes déterministes
- Algorithmes stochastiques
 - Algorithmes génétiques
 - Recuit simulé
 - Essaims particuliers
 - Recherche avec tabou

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

1 Algorithmes d'optimisation

- Algorithmes déterministes
- Algorithmes stochastiques
 - Algorithmes génétiques
 - Recuit simulé
 - Essaims particuliers
 - Recherche avec tabou

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Algorithmes génétiques – évolution différentielle (1/4)

Introduit par Storn et Price (1997) sous le nom d'algorithme à évolution différentielle.

- Paramètres dans un espace de dimension d .
- Algorithme agit sur N vecteurs de paramètres à partir de tirages aléatoires.
- Fonctionnement optimal pour $N \sim 5 \times d$.
- A chaque itération, chaque vecteur est modifié à partir des composantes des autres vecteurs.
- Recherche d'un minimum (ou maximum).
- Peu de conditions sur la fonction à optimiser (même pas besoin d'avoir la continuité).

Algorithmes génétiques – évolution différentielle (2/4)

On décrit algo après k itérations : on a N vecteurs de dimension d notés $x_{i,k}$, avec $i \in \{1, \dots, N\}$. On cherche $x_{i,k+1}$.

- ❶ **Mutation** : $\forall i \in \{1, \dots, N\}$, on définit un vecteur mutant comme combinaison de vecteurs de paramètres :

$$v_{i,k+1} = x_{r_1,k} + F(x_{r_2,k} - x_{r_3,k})$$

où les indices r sont tirés aléatoirement parmi $\{1, \dots, N\}$ et tels que i, r_1, r_2, r_3 sont tous différents, et F est le paramètre d'amplification de la différence.

- ❷ **Croisement** : $\forall i \in \{1, \dots, N\}$, le vecteur candidat $u_{i,k+1} = (u_{i,k+1}^1, u_{i,k+1}^2, \dots, u_{i,k+1}^d)$ hérite de composantes de $x_{i,k}$ et du vecteur mutant.

$$\forall j \in \{1, \dots, N\}, u_{i,k+1}^j = \begin{cases} v_{i,k+1}^j & \text{si } (U(j) \leq C) \text{ ou } j = I(i) \\ x_{i,k}^j & \text{si } (U(j) > C) \text{ ou } j \neq I(i) \end{cases}$$

où les $U(j)$ sont des tirages uniformes dans $[0, 1]$ et $I(i)$ un tirage uniforme dans $\{1, \dots, N\}$. C est la constante de croisement (*crossover*), qui est donc le paramètre d'une loi de Bernoulli.

- ❸ **Sélection** : La valeur de $x_{i,k+1}$ est donnée par un critère glouton : ce vecteur prend la valeur du vecteur le plus performant parmi $x_{i,k}$ et $u_{i,k+1}$.

Algorithmes génétiques – évolution différentielle (3/4)

Initialisation et convergence :

- La condition $j = I(i)$ permet d'avoir au moins une composante modifiée pour chaque vecteur (composante tirée au sort pour chaque vecteur).
- Initialisation : prendre N vecteurs bien écartés (tirés aléatoirement dans intervalle admissible ?) pour ne pas limiter la recherche localement.
- Critère d'arrêt : vecteurs dans une sphère de taille réduite : on conserve le meilleur des N vecteurs.
- Comment choisir la sphère limite pour éviter imprécision ou trop grand nombre d'itérations ? Lier aux données du problèmes. Pour calibrage de la vol implicite, choisir d'arrêter quand les prix du modèle sont à moins de 1 centime du prix observé, par exemple.
- La mutation est cohérent avec le besoin de convergence : quand les vecteurs sont proches, les mutants restent proches aussi (si F n'est pas trop grand).
- Malgré tout, le nombre d'itérations n'est pas facile à anticiper (on n'arrive pas à montrer que cet algorithme converge), donc on peut vouloir contrôler aussi ce nombre, indépendamment du critère d'arrêt.

Algorithmes génétiques – évolution différentielle (4/4)

Paramétrage, contraintes et minima locaux :

- F est lié à la vitesse de l'évolution et donc de la convergence. Il est conseillé de prendre F dans $[0.5, 1]$.
- C doit être grand mais différent de 1 (par exemple 0.95).
- Prise en compte de contraintes (par exemple paramètre de taux court d'un modèle de Nelson-Siegel doit être interprétable) : refaire le tirage du vecteur mutant tant qu'il ne vérifie pas les contraintes.
- Si la contrainte dépend de plusieurs paramètres (cas de Nelson-Siegel avec contrainte sur la somme de deux paramètres), il faut aussi contrôler le croisement.
- On accepte de s'éloigner du minimum pour une itération seulement (lors de la mutation).
- Une solution pour éviter le piège du minimum local : lancer plusieurs fois l'algorithme avec des valeurs initiales différentes.

1 Algorithmes d'optimisation

- Algorithmes déterministes
- Algorithmes stochastiques
 - Algorithmes génétiques
 - Recuit simulé
 - Essaims particuliers
 - Recherche avec tabou

2 Calibration de volatilité implicite

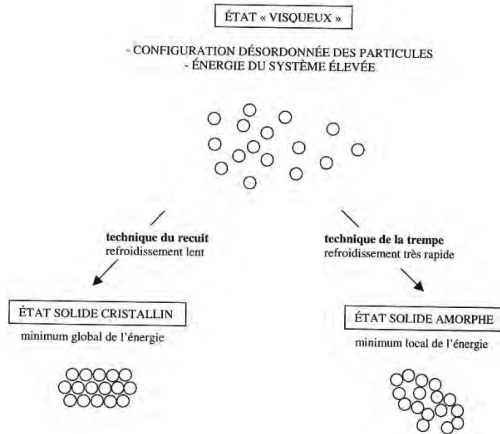
3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Recuit simulé

Recuit simulé (*simulated annealing*) inspiré du recuit en métallurgie : lors du refroidissement d'un métal fondu, après avoir un peu refroidi, on chauffe à nouveau puis on refroidit lentement, ce qui permet aux particules de trouver plus facilement leur état stable correspondant au minimum d'énergie.



Source : Jean-Philippe Préaux.

Recuit simulé

Principe du recuit simulé :

- Par analogie, on cherche à minimiser l'énergie E d'un système (fonction objectif).
- Pour un système dans un état donné x_k , on considère un autre état x_{k+1} :
 - Si $E_{k+1} \leq E_k$, on accepte ce nouvel état, meilleur (recherche locale de l'optimum).
 - Si $E_{k+1} > E_k$, on peut accepter ou refuser cet état. Si on accepte, on élargit la recherche. Choix par tirage de variable aléatoire paramétrée par la "température" T : plus T est élevé, plus on a tendance à accepter x_{k+1} .
- Aspect aléatoire permet de ne pas se laisser piéger par minimum local : on accepte ponctuellement de s'éloigner du minimum.
- La probabilité d'accepter x_{k+1} , état dégradant l'optimalité, est $e^{-(E_{k+1}-E_k)/k_B T}$. En physique statistique, c'est égal à p_{k+1}/p_k , où p_k est probabilité qu'une particule soit dans l'état E_k .

La **statistique de Maxwell-Boltzman**, en physique statistique donne la probabilité p_i d'une particule i dans un gaz (faibles interactions entre particules), à l'équilibre thermodynamique, d'être dans un état défini par son énergie E_i :

$$p_i = \frac{e^{-E_i/k_B T}}{\sum_j e^{-E_j/k_B T}},$$

où $k_B = 1.38 \times 10^{-23} \text{ J.K}^{-1}$ est la constante de Boltzmann.

Recuit simulé - Algorithme de Metropolis-Hastings

Le recuit simulé est une adaptation de l'algorithme de Metropolis-Hastings (1953), une méthode de Monte-Carlo par chaîne de Markov⁵ (MCMC).

- Soit x_i un vecteur (de grande dimension). On tire une variable x' , de même dimension depuis une distribution arbitraire et conditionnelle à x_i (prendre par exemple une distribution normale centrée en x_i), $q(x|x_i)$. La distribution q (*proposal density* ou *jumping distribution*) est choisie pour rendre le tirage de x' facile.
- La densité cible, dans laquelle on veut faire notre Mont-Carlo, est π : connue et compliquée de faire un tirage dedans.
- Soit le *ratio d'acceptation* $\alpha = \pi(x')/\pi(x_i)$.
 - Si $\alpha \geq 1$, alors $x_{i+1} = x'$: on privilégie nouvel état s'il est plus probable.
 - Si $\alpha < 1$, alors $x_{i+1} = x'$ avec probabilité⁶ α et $x_{i+1} = x_i$ avec probabilité $1 - \alpha$: on se laisse une possibilité d'accepter le nouvel état moins probable.
- Chaîne de Markov car x_{i+1} ne dépend que de x_i . Distribution des x_i converge vers π .

Lien du recuit simulé avec cet algorithme : à chaque valeur de la fonction objectif correspond une probabilité que ce soit l'optimum ; on tire des vecteurs de paramètres selon cette probabilité, i.e. en sur-pondérant les vecteurs dont l'image par la fonction objectif est faible.

5. Monte-Carlo s'appuyant sur le parcours d'une chaîne de Markov dont la loi stationnaire est la distribution à échantillonner.

6. La décision se fait en tirant une variable aléatoire uniforme u dans $[0, 1]$.

Recuit simulé

Pour notre problème d'optimisation :

- Fonction $f(x)$ à minimiser en fonction de x , vecteur de paramètres à calibrer.
- Analogie : x est l'état du système, $f(x)$ son énergie. La température est remplacée par un paramètre de contrôle, variable dans le temps, $c_k > 0$.
- Connaissant x_k , on tire aléatoirement x_{k+1} dans une distribution $q(x_{k+1}|x_k)$, par exemple une gaussienne centrée en x_k .
- La probabilité que x_{k+1} soit accepté est :
 - 1 si $f(x_{k+1}) \leq f(x_k)$;
 - $\exp\left(-\frac{f(x_{k+1})-f(x_k)}{c_k}\right)$ sinon.
- Si x_{k+1} est rejeté, on tire un nouveau x_{k+1} et on recommence test d'acceptation.
- Paramètre de contrôle initialement grand (grand taux d'acceptation) puis décroissance lente : $c_{k+1} = \alpha c_k$, avec $\alpha = 0.99$, par exemple.
- Fin de l'algorithme : x_k stationnaire (compter les tirages successifs rejetés).
- Défauts de la méthode : beaucoup de choix techniques (choisis empiriquement) : q , c_0 , dynamique de c_k (schéma de refroidissement), critère d'arrêt.
- Avantages : possibilité d'échapper au minimum local, désordre ajustable avec température.

1 Algorithmes d'optimisation

- Algorithmes déterministes
- Algorithmes stochastiques
 - Algorithmes génétiques
 - Recuit simulé
 - Essaims particuliers
 - Recherche avec tabou

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Essais particuliers

Optimisation par essaims particuliers (PSO : *particle swarm optimization*) s'inspire du comportement social d'oiseaux ou abeilles :

- Chaque particule représente un vecteur de paramètre.
- On part d'une population aléatoire dans l'espace de recherche.
- Chaque particule est caractérisée par sa position et son vecteur vitesse : elle adapte celui-ci en fonction de sa propre expérience (la meilleure position qu'elle a visitée) et de celle de ses voisines.
- Plus grande propension à explorer l'espace loin de l'optimum local (à cause de l'inertie et de la collaboration avec autres particules) qu'algorithmes génétiques.
- Minimisation (ou maximisation) d'une fonction f (appelée *fitness*).

Essais particuliers – algorithme

Pour chaque particule i , à étape k , on a :

- le vecteur position actuel, $x_{i,k}$, dans l'espace des paramètres à calibrer ;
- le vecteur vitesse actuelle, $v_{i,k}$;
- la meilleure position visitée, $P_{i,k} : f(P_{i,k}) = \min_{k' \leq k} f(x_{i,k'})$;
- la meilleure position trouvée par ses voisines, $G_{i,k} : f(G_{i,k}) = \min_{k' \leq k, i' \in I_{i,k}} f(x_{i',k'})$,
où $I_{i,k}$ est l'ensemble des voisins de i à l'instant k .

L'actualisation de x et v se fait ainsi :

$$\begin{cases} v_{i,k+1} &= \omega v_{i,k} + U_{i,k}(\phi_P)(P_{i,k} - x_{i,k}) + U_{i,k}(\phi_G)(G_{i,k} - x_{i,k}) \\ x_{i,k+1} &= x_{i,k} + v_{i,k+1}, \end{cases}$$

où

- ω est l'inertie ($1 - \omega$ agit comme un coefficient de frottement) ;
- $U_{i,k}(\phi)$ est une variable aléatoire uniforme tirée dans $[0, \phi]$;
- ϕ_P (resp. ϕ_G) est un coefficient d'accélération (ou de confiance) vers la meilleure position personnelle (resp. des voisins).

Les deux produits du type $U_{i,k}(\phi_P)(P_{i,k} - x_{i,k})$ peuvent être remplacés par des produits coordonnée par coordonnée si on prend un vecteur de variables aléatoires.

Essaims particulières – topologie du voisinage

- Intuitivement, topologie euclidienne dynamique : on calcule à chaque date les plus proches voisins selon distance euclidienne. En pratique ce n'est pas le meilleur choix (distances longues à déterminer, localise la recherche).
- Topologie statique *gbest* (global best) : toutes les particules sont connectées, on conserve uniquement le meilleur vecteur, commun à toutes les particules. Alors actualisation de $G_{i,k}$ par : si $\exists j, f(x_{j,k+1}) \leq f(G_{i,k})$, alors $G_{i,k+1} = x_{j,k+1}$, sinon $G_{i,k+1} = G_{i,k}$.
- Topologie statique *lbest* (local best) : chacune des N particules est liée à deux voisins (définir permutation σ de $[1, N]$ avec faible $\sum_i d(x_{\sigma(i),0}, x_{\sigma(i+1),0})$). Permet une recherche parallèle dans plusieurs lieux avec liens entre eux.
- *lbest* plus lent que *gbest* mais se laisse moins piéger par optima locaux : Suganthan (1999) propose une topologie dynamique commençant en *lbest* et finissant en *gbest*.
- Topologie de hiérarchie dynamique (Janson et Middendorf (2005)) : une hiérarchie (linéaire) est construite en fonction de la performance de chaque particule, $f(P_{i,k})$, et évolue avec le temps ; chaque particule est influencée par son supérieur hiérarchique direct.
- Clerc (2006) propose une topologie (TRIBES) basée sur des tribus, sous-essaims *gbest* indépendants qui évoluent dynamiquement : les bonnes tribus (avec $f(G_{i,k})$ faible) sacrifient leurs moins bons éléments, alors que les moins bonnes tribus recrutent de nouveaux membres (générés aléatoirement).

Essaims particulaires – convergence

- L'explosion de la vitesse peut être contrôlée par une contrainte supplémentaire : que la norme de la vitesse soit dans $[-v_{max}, +v_{max}]$, mais aussi par le contrôle de la position, limitée à une zone délimitée.
- Pour éviter ces contraintes, on a introduit $\omega \geq 0$, lié à stabilité de l'essaim : commencer près de 1 (instable si > 1) et diminuer progressivement (e.g. vers 0.4) pour assurer convergence (grande viscosité) ; ou prendre valeur constante vers 0.7.
- Des coefficients d'accélération trop grands conduisent à une divergence. Clerc et Kennedy (2002) recommandent de prendre $\phi_P + \phi_G \approx 4 \times \omega$ (avec $\phi_P = \phi_G$).
- Critère d'arrêt : nombre maximal d'itérations atteint, faibles variations de vitesse. La solution est donnée par la meilleure particule.

1 Algorithmes d'optimisation

- Algorithmes déterministes
- Algorithmes stochastiques
 - Algorithmes génétiques
 - Recuit simulé
 - Essaims particuliers
 - Recherche avec tabou

2 Calibration de volatilité implicite

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Recherche avec tabou

- Constat : beaucoup d'algorithmes passent un temps démesuré à affiner la recherche autour d'un optimum local et perdent ainsi du temps (et se laissent piéger dans le minimum local ou dans des cycles, incapable de sortir de l'optimum local).
- Solution : Créer une liste d'états prohibés, correspondant aux états déjà visités et conservés en mémoire (possibilité d'interdire le voisinage de ces états, pour un problème continu, la taille du voisinage pouvant diminuer progressivement afin de passer d'une recherche globale à une recherche locale)
- Utilisation
 - comme algorithme en tant que tel, recherche locale : partir d'un vecteur de paramètres initial x_i et itérer en ajoutant par exemple un tirage de variable aléatoire gaussienne ε_i si améliore optimisation ($f(x_i) > f(x_i + \varepsilon_i)$), ou en discrétisant (de plus en plus finement) l'espace des paramètres et en faisant des pas de taille 1 ; le tabou transforme la recherche local en autorisant à aller dans des états moins bons si on ne trouve pas d'état meilleur autorisé.
 - ou en hybridation (modification d'un autre algorithme).

Recherche avec tabou

Type de mémoire (les trois peuvent être utilisées ensemble ou séparément, la première étant la plus courante) :

- Court terme : durée de tabou δ fixée, interdiction de visiter les voisinages d'états visités il y a moins de δ itérations.
- Moyen terme : création de règles d'intensification des recherches dans une zone jugée prometteuse (on biaise la variable aléatoire permettant de définir le nouvel état, par exemple).
- Long terme : création de règles de diversification des recherches, avec un biais vers des zones encore inexplorées.

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
 - Modèle de Black-Scholes
 - Volatilité implicite
 - Volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Modèle de Black-Scholes – hypothèses

- Prix de l'actif sous-jacent modélisé par un mouvement brownien géométrique.
- Absence d'opportunités d'arbitrage et agents rationnels (marché efficient).
- Absence de coûts de transaction.
- Temps continu.
- Prix continus (prix de 98.53621 eur par exemple).
- Sous-jacent divisible (on peut acheter une portion d'action).
- Vente à découvert autorisée.
- Sans dividende et taux sans risque connu *ex ante* et constant (hypothèses contournables).

Modèle de Black-Scholes

Portefeuille de valeur V_t investi dans δ_t actions de prix S_t , alors que le reste épargné à taux r_t . Variation de valeur endogène :

$$dV_t = \delta_t dS_t + (V - \delta_t S_t) r_t dt.$$

V_t est une fonction de t et S_t : $V(t, S_t)$. Si S_t est un processus d'Ito, c'est-à-dire

$$dS_t = \mu_t dt + \sigma_t dW_t,$$

avec W_t un mouvement brownien (cas particulier : modèle de Black-Scholes avec $\mu_t = \mu S_t$ et $\sigma_t = \sigma S_t$, qui conduit à $S_t = S_0 \exp(\sigma B_t + \mu t - \frac{1}{2} \sigma^2 t)$), on peut appliquer le lemme d'Ito si V est $\mathcal{C}^2(\mathbb{R}_+ \times \mathbb{R}, \mathbb{R})$:

$$d(V(t, S_t)) = \frac{\partial V}{\partial t}(t, S_t) dt + \frac{\partial V}{\partial x}(t, S_t) dS_t + \frac{1}{2} \frac{\partial^2 V}{\partial x^2}(t, S_t) \sigma_t^2 dt.$$

Ainsi, le résultat important de Black, Scholes et Merton concerne la couverture du risque action, à savoir que $\delta_t = \frac{\partial V}{\partial x}(t, S_t)$: si on veut couvrir le risque action d'un portefeuille quelconque, il faut vendre une quantité $\frac{\partial V}{\partial x}(t, S_t)$ d'actions.

Modèle de Black-Scholes

Dans le cas particulier d'un call européen avec S_t un mouvement brownien géométrique et avec une série d'hypothèses (absence d'opportunité d'arbitrage, temps continu, ventes à découvert autorisées, pas de coûts de transaction, ni de dividendes et actifs parfaitement divisibles), on obtient le prix du call comme l'espérance du payoff actualisé sous proba risque neutre :

$$C(S_0, K, r, T, \sigma) = S_0 \mathcal{N}(d_1) - Ke^{-rT} \mathcal{N}(d_2),$$

avec $d_1 = \frac{1}{\sigma\sqrt{T}} \left[\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right) T \right]$ et $d_2 = d_1 - \sigma\sqrt{T}$. Les grecques (sensibilité aux facteurs de risque, qui peuvent être couverts par portefeuille répliquant où poids égal à ces grecques investi dans facteurs de risque, si disponibles) sont alors :

- $\delta = \partial C / \partial S = \mathcal{N}(d_1)$;
- $\Gamma = \partial^2 C / \partial S^2 = \mathcal{N}'(d_1) / S_0 \sigma \sqrt{T} > 0$;
- le vega $\mathcal{V} = \partial C / \partial \sigma = S_0 \sqrt{T} \mathcal{N}'(d_1) > 0$;
- etc. avec le thêta $\Theta = -\partial C / \partial T$ (donc dérivée sur vie restante), le rho $\rho = \partial C / \partial r$ et $\partial C / \partial K$.

En comparant les parties en dt de la formule d'Ito appliquée à C et de la formule de dV_t du portefeuille répliquant (donc $V_t = C_t$), on a l'EDP :

$$\frac{1}{2} \sigma^2 S_0^2 \Gamma + r S_0 \delta - r C + \Theta = 0.$$

Modèle de Black-Scholes – contexte et critiques

Un peu d'histoire :

- **17ème siècle** : utilisation d'options de vente et d'achat et vente à découvert (Amsterdam).
- **1973** : nouvelle méthode de valorisation et réplication des calls/puts avec la formule de Black-Scholes-Merton.
- **1973** : ouverture du Chicago Options Exchange (moins de 1000 options le premier jour, plus d'un million par jour en 1995).
- **1995** : faillite de la Barings à cause de straddles (short call et short put de même strike : prime à la stabilité) de Nick Leeson, révèle aux yeux de tous les effets de levier permis par les produits dérivés.

Critique de Mandelbrot :

- Distribution sous-jacente gaussienne irréaliste : minimisation des risques extrêmes.
Solution :
 - En tirer profit comme l'a fait Nassim Taleb en investissant sur les options très en-dehors de la monnaie.
 - Mettre une rustine au modèle mal spécifié : *smile* de volatilité.
- Hypothèses générales irréalistes : rationalité des investisseurs, pas d'opportunité d'arbitrage, etc. Solution : distribution alpha-stable ou mouvement brownien fractionnaire ?

Risque de modèle

Le risque de modèle peut provenir de plusieurs sources :

- le modèle utilisé est mal spécifié : les résidus sont distribués différemment de ce qui est supposé (utiliser modèles non-paramétriques, mais risque d'overfitting), un paramètre supplémentaire est nécessaire (mouvement brownien vs mouvement brownien fractionnaire), un facteur de risque n'est pas identifié ;
- il y a une incertitude dans l'estimation : présence de bruit (débruitage partiel), variation des paramètres dans le temps et impossibilité de prévoir leur valeur future ;
- même bien estimé, la sortie du modèle peut être incertaine : intervalles de confiance d'un Monte-Carlo, approximations numériques d'un schéma aux différences finies.

Modèle de Black-Scholes – risque de modèle

Stratégie de couverture du risque action d'une option dans le modèle de Black-Scholes avec une volatilité *fausse* $\bar{\sigma}$ au lieu de σ_t (volatilité changeant dans le temps, éventuellement aléatoire et martingale). On achète en 0 une option de payoff $h(S_T)$ et de prix C_t et on vend un portefeuille répliquant \bar{V}_t investi dans δ actions. Ce portefeuille évolue mais toujours en considérant la même volatilité, alors que l'on devrait prendre une volatilité dynamique. On dénoue en T . Profit est

$$e_T = \bar{V}_T - h(S_T) = \int_0^T e^{-r_t(T-t)} \frac{1}{2} (\bar{\sigma}^2 - \sigma_t^2) S_t^2 \Gamma_t dt.$$

Si le Γ est positif (c'est le cas pour les calls), cette erreur de couverture conduit à une perte si la volatilité est sous-estimée.

En effet, la variation de prix attendue du portefeuille en supposant une vol constante vaut, en utilisant la formule de réplication de Black-Scholes :

$$d(\bar{V}(t, S_t)) = \Theta_t dt + \delta_t dS_t + \frac{1}{2} \Gamma_t S_t^2 \bar{\sigma}^2 dt.$$

Au contraire, la "vraie" valeur du call vaut, par la formule d'Ito (ou Black-Scholes) :

$$d(C(t, S_t)) = \Theta_t dt + \delta_t dS_t + \frac{1}{2} \Gamma_t S_t^2 \sigma_t^2 dt.$$

Le point important est que les grecques sont identiques en t car $\sigma_t = \bar{\sigma}$ (mais évoluent différemment ensuite), ce qui permet d'écrire $de_t = \frac{1}{2} \Gamma_t S_t^2 (\bar{\sigma}^2 - \sigma_t^2) dt$.

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
 - Modèle de Black-Scholes
 - Volatilité implicite
 - Volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Volatilité implicite

Plusieurs notions de volatilité :

- La **volatilité instantanée** mesure l'intensité de la variation relative du prix du titre. Elle dépend du temps, elle peut être (modélisé de manière) aléatoire. C'est σ_t^X dans

$$\frac{dX_t}{X_t} = r_t dt + \sigma_t^X d\hat{W}_t.$$

Un call, dans le modèle de Black-Scholes, est instantanément plus volatil que le sous-jacent (effet de levier) : en effet

$$\frac{dC_t}{C_t} = r_t dt + \frac{\delta_t S_t}{C_t} \left(\frac{dS_t}{S_t} - r_t dt \right)$$

donc $Vol \left[\frac{dC_t}{C_t} \right] = \frac{\delta_t S_t}{C_t} Vol \left[\frac{dS_t}{S_t} \right]$, avec $C_t = \delta_t S_t - Ke^{-r(T-t)} \mathcal{N}(d_0) < \delta_t S_t$.

- La **volatilité historique** (mais la série est-elle stationnaire, l'hypothèse de log-normalité est-elle vérifiée ? En général non) : elle est souvent inférieure à la volatilité implicite.
- La **volatilité implicite** est le paramètre qui permet d'ajuster prix observé et prix stipulé par modèle : $C^{obs}(t, T, K) = C^{BS}(t, T, K, \Sigma^{impl})$.

Volatilité implicite

- En général, on observe un effet de smile, avec une volatilité implicite dépendant des paramètres de l'option, ce qui est contradictoire avec le modèle.
- La volatilité implicite ne fait référence à aucune notion statistique : on ne cherche pas le paramètre permettant de coller aux données en moyenne, mais on remplace le paramètre par un vecteur permettant de retrouver toutes les données.
- Volatilité implicite repose sur le fait que dans un **marché très liquide**, offre et demande s'équilibrent autour d'un consensus qui ne correspond pas à une erreur dans un sens statistique.
- Comment interpoler ce vecteur ?
- Volatilité implicite pourrait refléter la volatilité future anticipée par le marché (et pourtant en changeant les strikes, la volatilité implicite change), mais aussi toutes les incertitudes sur la qualité du modèle utilisé.
- Nappes de volatilité selon maturité résiduelle et strike.
- Forme de sourire de la volatilité implicite selon le strike : doutes sérieux sur la distribution des rendements (queue épaisse) ou argument économique (le titre de vol implicite plus élevée est plus risqué, peut-être à cause d'une liquidité moindre, d'un risque de modèle plus élevé [si le modèle est faux, l'erreur est plus sensible pour de grandes variations, donc pour des strikes éloignés du marché]).

Volatilité implicite

Voies pour coller davantage à la réalité des prix de marché.

- Malgré les limites du modèle, couverture de l'option nécessite la connaissance du delta et donc de la volatilité implicite.
- On peut estimer une volatilité implicite ou bien estimer une dynamique implicite sur la volatilité et non plus seulement un simple paramètre implicite (volatilité stochastique, volatilité locale...).
- On peut aussi définir un modèle plus complexe (plus réaliste) et en estimer les autres paramètres de manière statique (exposant de Hurst, par exemple).

Pour calculer cette volatilité implicite (elle est unique pour call/put), pas de formule fermée : on utilise des algorithmes d'optimisation.

Volatilité implicite

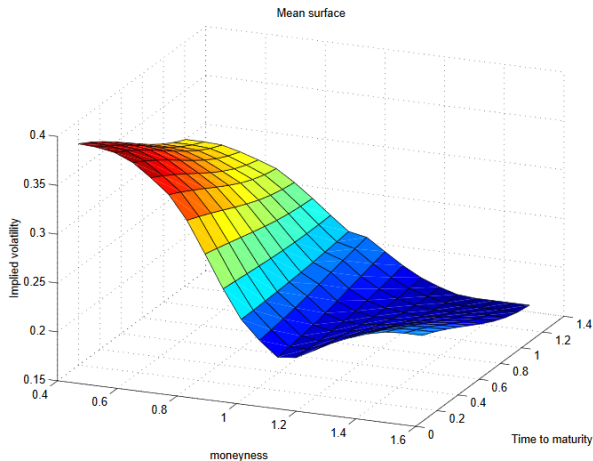


Fig. 1. Average implied volatility surface for SP500 options.

Source : Rama Cont, José da Fonseca.

Volatilité implicite – risque de modèle (1/3)

Une hypothèse pratique, plus très en vogue : **sticky** strike/moneyness/delta (à une date ultérieure, une option de même strike/moneyness (K/S) /delta conserve la même volatilité implicite).

En réalité, la nappe de volatilité évolue dans le temps. Voici une approche classique :

- 1 On identifie des facteurs dans la dynamique de la nappe (observée historiquement) ainsi que la matrice de corrélation les liant (une nappe est un champ aléatoire).
- 2 On estime une dynamique sur ces facteurs.
- 3 On projette ces facteurs à un horizon donné.
- 4 On déduit prévision de nappe à horizon donné.

Ou bien on peut faire plein de simulations avec les dynamiques de ces facteurs et, par Monte-Carlo, obtenir l'espérance de la future nappe.

Volatilité implicite – risque de modèle (2/3)

Risque de modèle classique pour un tel modèle de prévision :

- mauvais choix des facteurs, mauvaise spécification des dynamiques ou du lien linéaire supposé par la corrélation ;
- incertitude des méthodes d'estimation (nombre de données, loi des grands nombres) ;
- Difficulté à interpoler la nappe (pour un strike et maturité non répertoriés).

Pour ce dernier point, ne pas négliger la médiocrité de méthodes semblant simples... Illustration par le phénomène de Runge (interpolation polynomiale) :

Interpolation de Lagrange : fonction connue en n points $(x_1, y_1), \dots, (x_n, y_n)$. Un seul polynôme de degré au plus $n - 1$ passe par tous ces points (c'est du calibrage, pas de l'estimation !) :

$$L(X) = \sum_{i=1}^n y_i \prod_{j \neq i} \frac{X - x_j}{x_i - x_j}.$$

Plus n est grand, moins on peut contrôler ce qui se passe entre les x_i . En effet, le reste de la formule de Taylor appliquée à la fonction moins son polynôme de Lagrange est une dérivée d'ordre n de la fonction : très instable !

Volatilité implicite – risque de modèle (3/3)

Solutions contre le phénomène de Runge :

- Utiliser des x_i spécifiques : les abscisses de Tchebychev minimisent les fluctuations ; ce sont les racines des polynômes de Tchebychev $P_{n+1}(X) = 2XP_n(X) - P_{n-1}(X)$.
- Préférer des splines.
- Techniques de régularisation, par exemple : utiliser un polynôme de degré supérieur à celui du problème, mais en ajoutant une pénalité sur la régularité.

Volatilité implicite – Initialisation

Pour utiliser un algorithme d'optimisation, il faut d'abord l'initialiser. Cela peut être fait avec formule de Corrado et Miller (1996), s'appuyant sur une approximation linéaire de \mathcal{N} , obtenue par développement de Taylor en 0 (aussi appelé développement de McLaurin quand pris en 0) :

$$\mathcal{N}(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}}[z + \mathcal{O}(z^3)].$$

En utilisant l'approximation dans le prix du call :

$C(S_0, K, r, T, \sigma) = S_0 \mathcal{N}(d_1) - Ke^{-rT} \mathcal{N}(d_2)$, on obtient la formule d'approximation suivante pour la volatilité implicite, en posant $X = Ke^{-rT}$:

$$\sigma^{impl} \sqrt{T} \approx \sqrt{2\pi} \left(\frac{C - (S - X)/2}{S + X} \right) + \sqrt{2\pi \left(\frac{C - (S - X)/2}{S + X} \right)^2 - 4 \left(\frac{S - X}{S + X} \right)^2}.$$

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
 - Modèle de Black-Scholes
 - Volatilité implicite
 - Volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Volatilité locale

Une possibilité pour avoir un modèle cohérent avec prix des calls (ie pour expliquer le smile plus que pour le supprimer) : volatilité locale, ie non constante, fonction déterministe de t et S_t :

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma(t, S_t) dW_t.$$

- Exemple de volatilité locale, le CEV (constant elasticity of variance, John Cox, 1975) : $dS_t = \mu S_t dt + \sigma_0 S_t^\gamma dW_t$, avec $\gamma \geq 0$. Si $\gamma < 1$, on observe l'effet de levier propre aux actions (vol augmente quand prix diminue) ; si $\gamma > 1$, effet de levier inverse propre aux commodities.
- Pour t et S_t donnés, selon le modèle de volatilité locale :
 - pour toutes les options on a la même volatilité locale ;
 - si on supposait une vol constante (ne dépend ni de t ni de S_t) comme dans le modèle de Black-Scholes, il faudrait s'attendre à avoir une volatilité implicite différente selon K et T , car les volatilité $\sigma(t', S_{t'})$, pour $t' > t$, ne contribuent pas de la même manière au prix (par exemple, la volatilité locale pour S faible a peu d'importance pour un call hors de la monnaie, contrairement au call dans la monnaie) : ce modèle de volatilité locale justifie l'existence du smile dans l'approche BS.

Volatilité locale

- Formule de Dupire (1994), donnant la volatilité locale implicite pour calls européens, dans cas particulier où $t = T$ et où $S_t = K$:

$$\sigma(T, K) = \sqrt{\frac{2 \left(\frac{\partial C}{\partial T} + rK \frac{\partial C}{\partial K} \right)}{K^2 \frac{\partial^2 C}{\partial K^2}}}.$$

- Nécessite continuum de strikes et maturités (et surface peu stable dans le temps, pas facilement utilisable). Modèle non-paramétrique, nécessite régularisation avant différenciation et interpolation. Il existe des modèles paramétriques, permettant en plus de vérifier qu'il n'y a pas d'arbitrages dans la surface de volatilité :
 - Modèle de Schonbucher (1999), à quatre paramètres (a, b, σ, ν), cohérent avec modèle d'Heston, pour $x = \ln(K/F_T)$, log-moneyness, où $F = S_0 \int_0^T r_t dt$ prix forward :

$$\sigma_{imp}^2(x) = a + bx + \sqrt{\sigma^4/4 + x^2\nu^2}.$$

- Modèle SVI (stochastic volatility inspired), aussi cohérent avec Heston, fonction aussi de log-moneyness :

$$\sigma_{imp}^2(x) = a + b \left[\rho(x - m) + \sqrt{(x - m)^2 + \sigma^2} \right].$$

- Théorème de Dupire : tout modèle à volatilité stochastique peut s'écrire sous la forme d'un modèle à volatilité locale.

Volatilité locale – SVI (1/2)

Figure 1: Impact of a change in ρ from 0 to -0.8 in the SVI/gSVI model ($b = 0.35$, $a = 0.01$, $\sigma = 0.01$, and $m = 0$ stay constant)

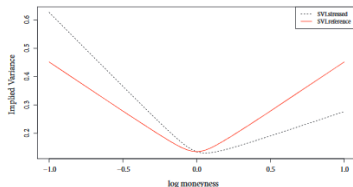


Figure 2: Impact of a change in b from 0.35 to 0.5 in the SVI/gSVI model ($\rho = 0$, $a = 0.1$, $\sigma = 0.01$, and $m = 0$ stay constant)

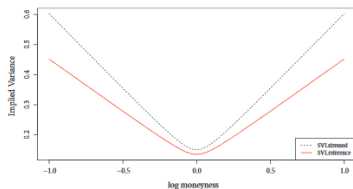


Figure 4: Impact of a change in σ from 0.01 to 0.03 in the SVI/gSVI model ($\rho = 0$, $b = 0.35$, $a = 0.1$, and $m = 0$ stay constant)

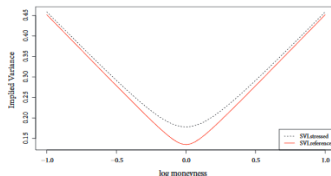
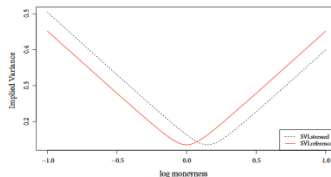


Figure 5: Impact of a change in m from 0 to 0.15 in the SVI/gSVI model ($\rho = 0$, $b = 0.35$, $a = 0.1$, and $m = 0$ stay constant)



Source : Damghani et Kos.

Volatilité locale – SVI (2/2)

Figure 3: Impact of a change in α from 0.1 to 0.15 in the SVI/gSVI model ($\rho = 0$, $b = 0.35$, $\sigma = 0.01$, and $m = 0$ stay constant)

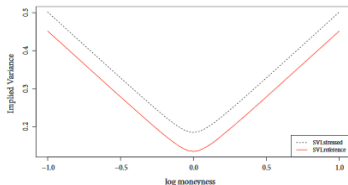
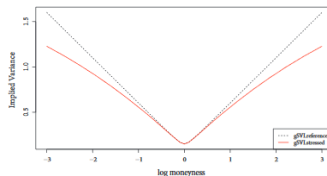


Figure 6: Impact of a change in β from 1 to 1.1 in the gSVI model ($\rho = 0$, $b = 0.35$, $\alpha = 0.1$, $\sigma = 0.01$, and $m = 0$ stays constant)



Source : Damghani et Kos.

The generalized SVI (gSVI) applies the SVI model to $z = (x - m)/\beta^{|x-m|}$ instead of x , with $\beta \in [1, 1.4]$, what allows having z an increasing function of the moneyness x , for x between 0 and 3. The gSVI bypasses one limitation of the SVI for out-of-the-money options, which suffers from the linear asymptote of the volatility in the SVI approach.

Volatilité locale – valorisation d'options

On n'a plus de formule fermée, mais on peut utiliser un Monte-Carlo ou l'EDP (schéma aux différences finies ou arbre binomial)

$$r_t C(t, S_t) = \partial_t C + r_t S_t \partial_S C + \frac{1}{2} \sigma(t, S_t)^2 S_t^2 \partial_S^2 C$$

avec condition terminale :

$$C(T, S) = (S_T - K)^+.$$

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Volatilité stochastique

Un facteur de risque additionnel par rapport à BS : la volatilité. Equations génériques :

$$\begin{cases} \frac{dS_t}{S_t} = rdt + \sigma_t dW_t^1 \\ \sigma_t = f(U_t) \\ dU_t = (\lambda_1 + \lambda_2 U_t)dt + (\nu_1 + \nu_2 U_t)^\alpha dW_t^2 \\ \langle dW_t^1, dW_t^2 \rangle = \rho dt. \end{cases}$$

Cas particuliers :

- modèle d'Heston, $\alpha = 1/2$, $\nu_1 = 0$, $f : x \mapsto \sqrt{x}$;
- modèle SABR, $\lambda_1 = \nu_1 = 0$, $\alpha = 1$, $f : x \mapsto x(S_t)^\beta$ (mélange de volatilité stochastique et de volatilité locale) ;
- modèle CEV (volatilité locale), $\lambda_1 = \lambda_2 = \nu_1 = \nu_2 = 0$ et $f : x \mapsto \sigma_0(S_t)^{\gamma-1}$.

Approches alternatives possibles comme le modèle MSM (Markov-switching multifractal).

Volatilité stochastique

Dans un modèle à volatilité stochastique, une couverture uniquement sur l'actif risqué ne fonctionne pas. Il faut aussi se couvrir contre les variations de volatilité, deuxième facteur de risque. Pour que le marché soit complet (=risque répliquable) on suppose qu'on peut donc investir dans actif liquide (un variance swap ou un autre produit dérivé sur S_t et σ_t) de prix $M_t = M(t, S_t, \sigma_t)$ tel que $\partial M / \partial \sigma \geq 0$. La valeur du portefeuille de réplication est :

$$dV_t = \delta_t dS_t + \omega_t dM_t + (V - \delta_t S_t - \omega_t M_t) r_t dt.$$

Le delta-vega hedging se fait donc avec

$$\delta_t = \frac{\partial C}{\partial S} - \omega_t \frac{\partial M}{\partial S}$$

et

$$\omega_t = \frac{\partial C / \partial \sigma}{\partial M / \partial \sigma}.$$

Volatilité stochastique

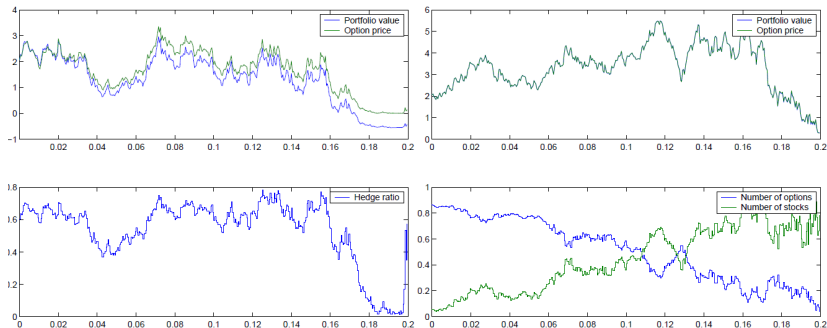


Figure 4.3: Couverture en vega. Gauche: couverture en delta seulement dans un marché à volatilité stochastique. Droite: couverture en delta-vega.

FIGURE — Source : Peter Tankov.

Volatilité stochastique – modèle d'Heston

Modèle d'Heston est un modèle avec retour à la moyenne : $\sigma_t = \sqrt{U_t}$ avec U_t la variance des rendements du sous-jacent, qui suit un modèle similaire au modèle CIR (Cox-Ingersoll-Ross, utilisé pour les taux) :

$$dU_t = \kappa(\theta - U_t)dt + \xi\sqrt{U_t}dW_t^2.$$

- $\theta > 0$ est la variance de long terme ;
- $\kappa > 0$ est la force de retour à la moyenne ;
- ξ est la vol de la vol.

ξ joue sur la kurtosis du smile, ρ sur sa skewness.

Si on avait 1 au lieu de $\sqrt{U_t}$, comme dans le modèle de Vasicek (qui est un Ornstein-Uhlenbeck appliqué aux taux), on ne serait pas assuré de la positivité de U_t . Ici, avec la condition de Feller (rarement vérifiée après calibration sur le marché) ($2\kappa\theta > \xi^2$), alors U_t est toujours strictement positif (sans cette condition, 0 peut être atteint en théorie, ou même dépassé en pratique si discrétisation pas assez fine du temps).

Volatilité stochastique – modèle d'Heston

Propriétés du modèle d'Heston :

- Volatilité de U level-dependent $\xi\sqrt{U_t}$ (raison pour laquelle on ne va pas atteindre une volatilité négative).
- Processus U stationnaire, de distribution gamma : la densité est

$$\delta(u_t) = \frac{\omega^\nu}{\Gamma(\nu)} u_t^{\nu-1} e^{-\omega u_t},$$

où $\omega = 2\kappa/\xi^2$ et $\nu = 2\kappa\theta/\xi^2$.

- Incréments non-stationnaires. Densité de U_{t+T} conditionnellement à U_t a formule explicite aussi, faisant intervenir une loi χ^2 . En particulier :

$$\mathbb{E}[u_{t+T}|u_t] = u_t e^{-\kappa T} + \theta(1 - e^{-\kappa T}) \xrightarrow{T \rightarrow +\infty} \theta,$$

$$\text{Var}[u_{t+T}|u_t] = u_t \frac{\xi^2}{\kappa} \left(e^{-\kappa T} - e^{-2\kappa T} \right) + \frac{\theta \xi^2}{2\kappa} \left(1 - e^{-\kappa T} \right)^2 \xrightarrow{T \rightarrow +\infty} \frac{\theta \xi^2}{2\kappa}.$$

Volatilité stochastique – modèle d'Heston

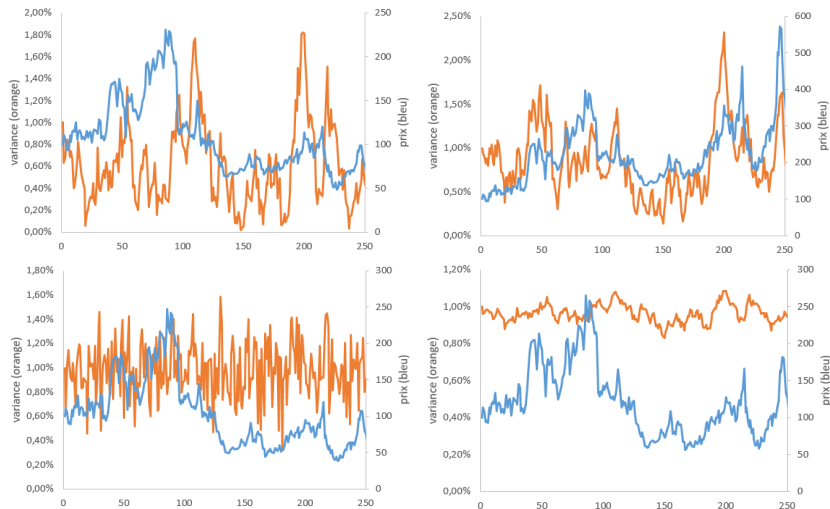


FIGURE — Simulations pour $\kappa = 5\%$, $\theta = 1\%$, $\xi = 2\%$, $\rho = -20\%$ (haut gauche, cas de base), puis $\rho = 50\%$ (haut droite), $\kappa = 100\%$ (bas gauche) et $\xi = 0.2\%$ (bas droite).

Volatilité stochastique – modèle SABR

SABR : stochastic alpha bêta rho.

$$\left\{ \begin{array}{l} \frac{dS_t}{S_t} = rdt + \sigma_t S_t^\beta dW_t^1 \\ d\sigma_t = \alpha \sigma_t dW_t^2 \\ \langle dW_t^1, dW_t^2 \rangle = \rho dt. \end{array} \right.$$

- $\beta \in [0, 1]$.
- On note v la vol initiale σ_0 . Alors on a une formule explicite pour la nappe de volatilité implicite, qui dépend de v (et de $F_t = S_t e^{-rt}$, prix forward).
- Le problème réside dans l'estimation de v . Recours possible à estimateur historique, puis calibrage des autres paramètres.

Volatilité stochastique – modèle SABR

Theorem 6.4 (volatilité implicite dans le modèle de SABR). *Soit le modèle de SABR $\ln(F/K)$ et $\tau = T - t$. Définissons les variables*

Soit $x =$

$$u = \frac{\alpha x}{v}$$

et

$$z = \frac{\alpha}{v} \frac{S_0^{1-\beta} - K^{1-\beta}}{1-\beta}.$$

Nous avons la relation suivante:

$$I_{\text{imp}}(x, \tau) = I^0(x) [1 + I^1(x)\tau] + \mathcal{O}(\tau^2)$$

avec:

$$I^0(x) = \begin{cases} vK^{\beta-1}, & \text{si } x = 0, \\ \frac{xv(1-\beta)}{S_0^{1-\beta} - K^{1-\beta}}, & \text{si } \alpha = 0, \\ \frac{\alpha x}{\ln\left(\frac{\sqrt{1-2\rho u + u^2} + u - \rho}{1-\rho}\right)}, & \text{si } \beta = 1, \\ \frac{\alpha x}{\ln\left(\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right)}, & \text{si } \beta \leq 1. \end{cases}$$

et

$$I^1(x) = \frac{(\beta-1)^2}{24} \frac{v^2}{(S_0 K)^{1-\beta}} + \frac{1}{4} \frac{\rho \alpha \beta v}{(S_0 K)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24} \alpha^2$$

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Densité risque-neutre

Possibilité d'avoir de nombreuses statistiques risque-neutre (ie pour lesquelles prix de modèle = prix de marché) :

- moments implicites ;
- paramètres de queue implicites ;
- mesures de risque implicite.

Toutes en fait sont des métriques liées à la densité implicite. Celle-ci peut-être générée de plusieurs manières :

- de manière paramétrique (il suffit de calibrer les paramètres) ;
- de manière non-paramétrique avec moments implicites et fonction génératrice des moments : $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{i=0}^{\infty} \frac{t^i m_i}{i!}$, par développement de l'exponentielle ;
- de manière non-paramétrique directe (à partir du prix des options) avec la formule de Breeden-Litzenberger.

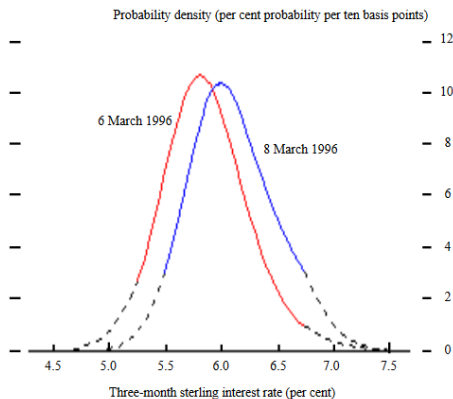
Objectif de ces statistiques risque-neutre : leur trouver un pouvoir prédictif, trouver des mesures de risque répondant à l'anticipation des marchés (qui de fait sont sensibles aux variations historiques et ne permettent donc pas d'avoir des mesures de risque robustes), affiner pricing et réplication.

Densité risque-neutre – Breeden-Litzenberger

- Calibrate all the density on options : $q(S_T) = e^{r(T-t)} \frac{\partial^2 C(t, S_t, K, T)}{\partial K^2} \Big|_{K=S_T}$.
- This distribution is calibrated at a single date, not on historical data : it is therefore much more responsive to changing market expectations.
- Justification de l'équation :
 - butterfly : achat de deux calls de strike $S_0 + \Delta S$ et $S_0 - \Delta S$ et vente de deux calls de strike S_0 .
 - Payoff $h(S_t)$ est triangle centré en S_0 , gain maximal ΔS , donc le payoff divisé par ΔS tend vers une fonction de Dirac quand $\Delta S \rightarrow 0$. Ainsi $h(S_T)/\Delta S \xrightarrow{\Delta S \rightarrow 0} \mathbb{1}_{S_T=K}$.
 - En passant à l'espérance, on retrouve prix des calls fois $e^{r(T-t)}$ à gauche et probabilité à droite (ie densité de probabilité fois dS).
- Besoin de connaître un continuum de prix et problème inverse mal posé. On peut le faire en interpolant de manière lisse prix ou (mieux car en général plus régulier) en interpolant de manière lisse les volatilités implicites (technique de Shimko).
- Difficulté malgré tout à représenter les queues de distributions (car extrapolation plus délicate qu'interpolation) : préférer approche paramétrique pour la queue ?

Densité risque-neutre – Breeden-Litzenberger

Implied RND functions for the three-month sterling interest rate in June 1996

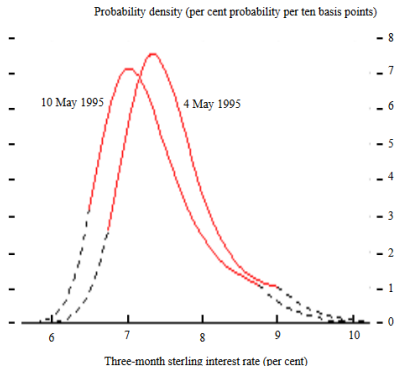


- (a) Derived using LIFFE June 1996 options on the short sterling future, as at 6 March and 8 March 1996. These options expire on 19 June 1996. These graphs illustrate the result when using the Shimko technique. In this case the smile curve was interpolated using a cubic spline with two knot points. The tails, which are shown as dotted lines, are lognormal.

FIGURE — Source : Bhupinder Bahra.

Densité risque-neutre – Breeden-Litzenberger

Implied RND functions for the three-month sterling interest rate in September 1995



- (a) Derived using LIFFE September 1995 options on the short sterling future, as at 4 May and 10 May 1995. These options expire on 20 September 1995. These graphs illustrate the result when using the Shimko technique. In this case the smile curve was interpolated using a cubic spline with two knot points. The tails, which are shown as dotted lines, are lognormal. Note that this extrapolation procedure does not always result in a smooth transition from the observable part of the distribution to the tails, as evidenced by the kink in the right tails.

FIGURE — Source : Bhupinder Bahra.

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire
 - Le mouvement brownien fractionnaire
 - Calibration de Black-Scholes fractionnaire
 - Au delà du mbf

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Fractales

Mandelbrot est le premier à en montrer l'omniprésence dans notre monde (*The fractal geometry of nature*, 1982), en particulier dans les cours de bourse.

Définition

Un ensemble $S \subset \mathbb{R}^n$ est dit *autosimilaire* s'il est la réunion de sous-ensembles disjoints obtenus par dilatation, rotation et translation de S .

Exemples

- flocon de Koch ;
- ensemble triadique de Cantor (on part de $[0, 1]$, on divise en trois intervalles en on enlève l'intervalle médian ouvert, $]1/3, 2/3[$, etc.).

Conséquence

Les fractales contiennent une infinité de singularités (comme les sommets du triangle de base du flocon de Koch ; mais, en finance, ces singularités peuvent être des *breaks* structurels, des sauts, etc.), dont on peut même rechercher la distribution. L'outil idéal pour les repérer : les ondelettes !

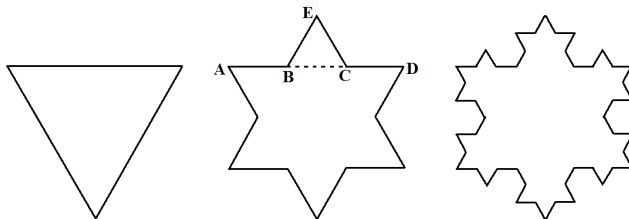


FIGURE – Flocon de Koch.

Dimension fractale

La surface du flocon de Koch est contenue dans un ensemble borné de \mathbb{R}^2 . Quelle est sa longueur ?

Un objet de longueur infinie et d'aire finie devrait donc avoir une dimension $d \in]1, 2]$.

L'utilisation de la dimension fractale (en fait dimension de capacité, de Minkowski-Bouligand ou de *box counting*, parfois égale (jamais inférieure, l'égalité a lieu en cas de stricte autosimilarité) à la dimension de Hausdorff [1919], la *vraie* dimension fractale, qu'elle simplifie) permet de définir une métrique invariante selon l'échelle.

Si $d \in \mathbb{N}$

Pour commencer, qu'est-ce qu'une dimension entière ? Considérons $\mathcal{S} = [0, 1] \times [0, 1]$. Il faut $N(s) = s^{-2}$ boules (au sens de la norme infinie dans \mathbb{R}^2) de diamètre s pour recouvrir \mathcal{S} , cela quel que soit s . Ainsi, $N(s)$ est proportionnel à s^{-2} et 2 est la dimension de \mathcal{S} .

Dimension fractale

Soit S un sous-ensemble borné de \mathbb{R}^n .

Définition

La **dimension fractale** de S est d tel que $N(s)$ est proportionnel à s^{-d} lorsque $s \rightarrow 0$.
Donc

$$d = -\lim_{s \rightarrow 0} \frac{\log N(s)}{\log s},$$

Généralisé à des fins de facilité d'estimation à

$$d = -\liminf_{s \rightarrow 0} \frac{\log N(s)}{\log s}.$$

Définition

La **mesure** de S (la distance en dimension 1, l'aire en dimension 2) est alors définie par $\limsup_{s \rightarrow 0} N(s)s^d$.

Dimension fractale

Exemples :

- flocon de Koch : $N(s) = 4^n$ pour $s = 3^{-n}$, donc $d = \log 4 / \log 3 = 1.26... \in]1, 2[$;
- ensemble triadique de Cantor : $N(s) = 2^n$ pour $s = 3^{-n}$, donc $d = \log 2 / \log 3 = 0.63... \in]0, 1[$;
- surface pulmonaire : $d = 2.97$;
- l'ensemble des zéros d'un mouvement brownien : $d = 0.5$;
- le graphe d'un mouvement brownien : $d = 1.5$ (*quid* du calcul de l'intégrale de la variation quadratique ?);
- Dow Jones, côte de Norvège (1.52) (Feder, *Fractals* (1988)), côte de Grande-Bretagne (1.24);
- boule de papier froissé : $d = 2.4 \pm 0.2$;
- Les séries que nous avons présentées ci-avant en HF.

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire
 - Le mouvement brownien fractionnaire
 - Calibration de Black-Scholes fractionnaire
 - Au delà du mbf

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Mouvement brownien fractionnaire



FIGURE – Hommage de Google à Mandelbrot le 20/11/20

Définition

Un **mouvement brownien fractionnaire** B_H d'exposant de Hurst $0 < H < 1$ est un processus gaussien tel que $B_H(0) = 0$, $\mathbb{E}\{B_H(t)\} = 0$ et

$$\mathbb{E}\{(B_H(t) - B_H(s))^2\} = \sigma^2 |t - s|^{2H}.$$

Définition intégrale :

$$B_H(t) = \frac{1}{\Gamma(H + \frac{1}{2})} \int_{-\infty}^{\infty} \left((t-s)_+^{H-1/2} - (-s)_+^{H-1/2} \right) dW_s.$$

Mouvement brownien fractionnaire

Proposition

- $|B_H(t) - B_H(s)|$ est proportionnel à $|t - s|^H$, ce qui correspond à la définition d'une fonction singulière partout (c'est donc une fractale homogène) d'exposant de Hölder (uniforme) H .
- Le MBf est statistiquement autosimilaire puisque $B_H(t)$ et $c^{-H}B_H(ct)$ ont la même distribution.^a
- Le MBf n'est pas stationnaire car sa covariance ne dépend pas que de $(t - s)$:

$$\mathbb{E}\{B_H(t)B_H(s)\} = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

- Les incréments du MBf sont stationnaires : $B_H(t) - B_H(s)$ a même distribution que $B_H(t - s) - B_H(0)$.
- Avec probabilité 1, le graphe de B_H a une dimension fractale de $2 - H$.

a. Ce sont des gaussiennes, et on montre qu'elles ont même espérance (0) et même variance.

Mouvement brownien fractionnaire

Un mouvement brownien classique est d'exposant de Hurst $H = 1/2$. Plus H est faible, plus le MBf est singulier. Le MBf a des propriétés dépendant de la valeur de H :

- Si $H = 1/2$, le MBf est un mouvement brownien classique, incréments **indépendants**.
- Si $H > 1/2$, les incréments sont **positivement corrélés** (efface singularités) ; c'est une primitive fractionnaire du mouvement brownien classique.
- Si $H < 1/2$, les incréments sont **négativement corrélés** (crée singularités) ; c'est une dérivée fractionnaire du mouvement brownien classique.

$$\begin{aligned}
 & \text{Correl}(B_H(n+1) - B_H(n), B_H(1) - B_H(0)) \\
 = & \frac{\mathbb{E}\{(B_H(n+1) - B_H(n))B_H(1)\}}{\sqrt{\mathbb{E}\{(B_H(n+1) - B_H(n))^2\}\mathbb{E}\{B_H(1)^2\}}} \\
 = & \frac{[\mathbb{E}\{B_H(n+1)B_H(1)\} - \mathbb{E}\{B_H(n)B_H(1)\}]/\sigma^2}{[(n+1)^{2H} - 2n^{2H} + (n-1)^{2H}]/2} \\
 \stackrel{\text{dév. lim.}}{=} & \frac{H(2H-1)}{n^{2(1-H)}} + o_{n \rightarrow \infty} \left(\frac{1}{n^{2(1-H)}} \right).
 \end{aligned}$$

Mouvement brownien fractionnaire

- L'autocorrélation décroît d'autant plus vite avec n que H est faible.
- L'autocorrélation pour $H > 1/2$ est plus forte à mesure que H monte.
- Pour $H < 1/2$, le H le plus faible n'implique pas l'autocorrélation la plus forte en valeur absolue. Cela dépend de n .
- Il y a une longue dépendance pour $H > 1/2$, ce qui signifie que

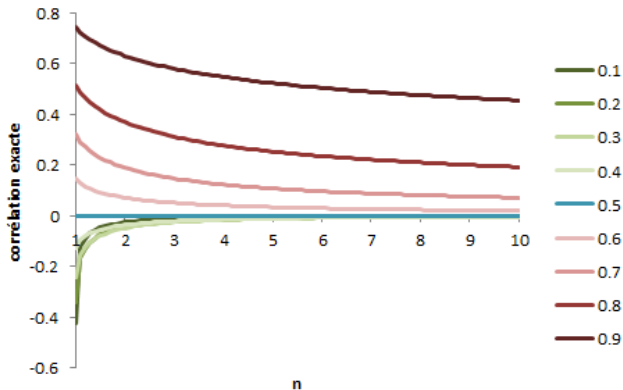
$$\sum_{n=1}^{\infty} \text{Correl}(B_H(n+1) - B_H(n), B_H(1) - B_H(0)) = +\infty :$$

l'autocorrélation devient proche de zéro quand n est grand, mais pas négligeable pour autant.

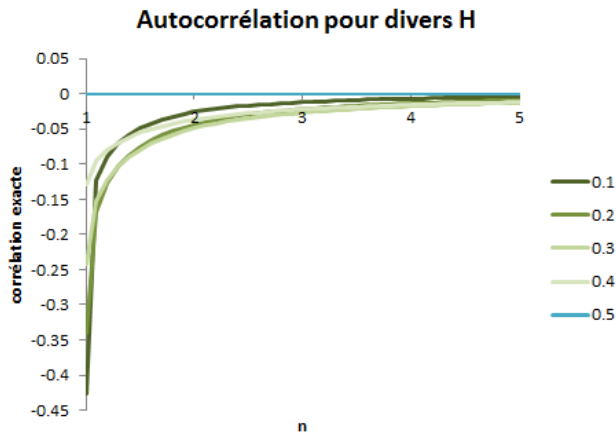
$$\begin{aligned} & \sum_{n=1}^N \text{Correl}(B_H(n+1) - B_H(n), B_H(1) - B_H(0)) \\ = & \frac{1}{2} \sum_{n=2}^{N+1} n^{2H} - \sum_{n=1}^{N+1} n^{2H} + \frac{1}{2} \sum_{n=0}^{N-1} n^{2H} \\ \sim & -\frac{1}{2} + HN^{2H-1} \xrightarrow{N \rightarrow +\infty} +\infty. \end{aligned}$$

Mouvement brownien fractionnaire

Autocorrélation pour divers H



Mouvement brownien fractionnaire



Mouvement brownien fractionnaire - Estimation

$$\mathbb{E}\{(B_H(t) - B_H(s))^2\} = \sigma^2 |t - s|^{2H},$$

donc, si on observe le signal sur $[0, 1]$, pour une discrétisation $(0/N, 1/N, 2/N, \dots, N/N)$, $B_H((i+1)/N) - B_H(i/N)$ a pour variance empirique une valeur tendant vers $\sigma^2 N^{-2H}$.

Donc :

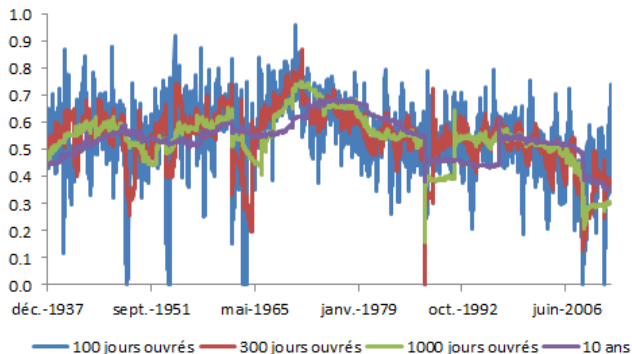
$$\hat{H} = \frac{\log(\sigma^2) - \log\left(\frac{1}{N} \sum_{i=0}^{N-1} (B_H((i+1)/N) - B_H(i/N))^2\right)}{2 \log(N)} \rightarrow_{p.s.} H.$$

Si σ est inconnu, il faut estimer H et σ . Pour cela, soit on estime d'abord σ , qui est la vol un jour, soit on fait le même calcul avec un pas de temps de $2/N$ au lieu de $1/N$ et on obtient cet autre estimateur de H (ici pour N pair) :

$$\hat{H} = -\frac{1}{2 \log(2)} \log \left(\frac{\frac{1}{2} \sum_{i=0}^{N-1} (B_H((i+1)/N) - B_H(i/N))^2}{\sum_{i=0}^{(N/2)-1} (B_H(2(i+1)/N) - B_H(2i/N))^2} \right).$$

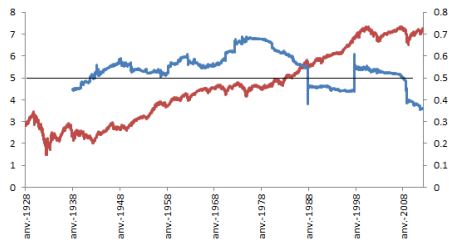
Mouvement brownien fractionnaire - Estimation

Exposant de Hurst, plusieurs tailles de fenêtres

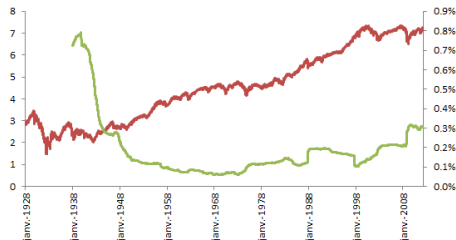


Estimation pour le logarithme népérien du prix pour S&P500, avec fenêtre glissante 10 ans, comme volatilité. Les deux pics sont le fait de la crise du 19/10/1987 (pire journée sur un marché action après krach de la bourse islandaise en 2008, et qui est à l'origine de l'instauration à Wall Street de coupe-circuits pour éviter les baisses trop importantes).

Mouvement brownien fractionnaire - Estimation



— ln-prix (axe gauche) — exposant de Hurst (axe droit)



— ln-prix (axe gauche) — volatilité 1 jour annualisée (axe droit)

Mouvement brownien fractionnaire - Estimation

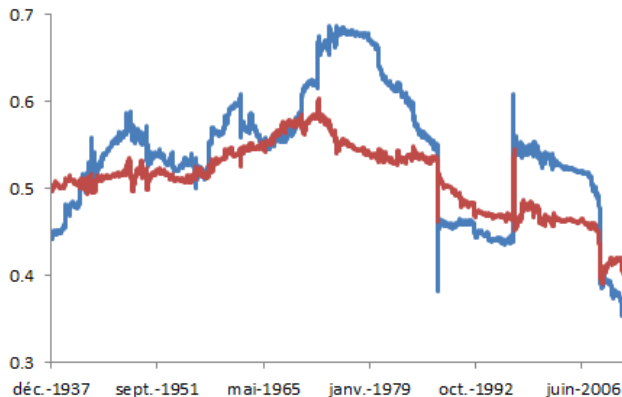


FIGURE – Estimateur avec pas de 2 jours et 1 jour (en bleu) vs estimateur avec pas de 10 jours et 1 jour (en rouge) : changer ces pas de temps permet de comparer directement plusieurs échelles et fournit donc des informations complémentaires.

Mouvement brownien fractionnaire - Simulation

- 1 Soit Γ la matrice de covariance définie par $\Gamma_{ij} = (i^{2H} + j^{2H} - |i - j|^{2H})\sigma^2/2$.
- 2 Obtenir Σ , la matrice "écart type", i.e. telle que $\Sigma^2 = \Gamma$ par décomposition de Choleski (André-Louis Choleski : si A est symétrique définie positive, il existe L triangulaire inférieure telle que $A = LL^t$).
- 3 Construire un vecteur X de variables aléatoires gaussiennes standard.
- 4 $Y = \Sigma X$ est un tirage de MBf.

Si $H = 1/2$, alors :

$$\Gamma = \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \dots \\ \sigma^2 & 2\sigma^2 & 2\sigma^2 & \dots \\ \sigma^2 & 2\sigma^2 & 3\sigma^2 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} \sigma & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \sigma & \dots & \dots & \sigma \end{pmatrix}$$

Mouvement brownien fractionnaire - Simulation

Décomposition de Choleski, pour $A = LL^t$, on calcule dans l'ordre des équations suivantes, la première colonne :

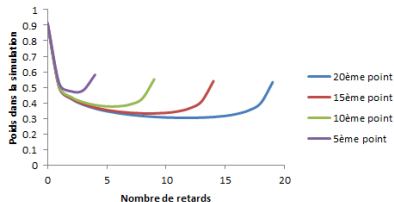
$$\begin{cases} l_{11} = \sqrt{a_{11}} \\ l_{j1} = a_{1j} / l_{11} \end{cases} \quad \text{pour } j \geq 2$$

puis la colonne $i \geq 2$:

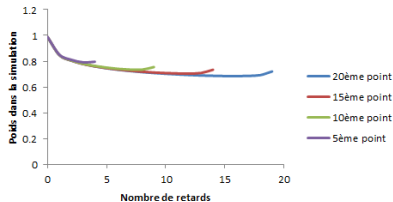
$$\begin{cases} l_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2} \\ l_{ji} = \frac{a_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk}}{l_{ii}} \end{cases} \quad \text{pour } j \geq i + 1.$$

Mouvement brownien fractionnaire - Simulation

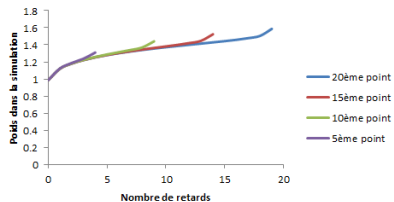
Poids dans la matrice de Choleski pour chaque v.a.
gaussienne standard, $H=0.2$



Poids dans la matrice de Choleski pour chaque v.a.
gaussienne standard, $H=0.4$



Poids dans la matrice de Choleski pour chaque v.a.
gaussienne standard, $H=0.6$



Poids dans la matrice de Choleski pour chaque v.a.
gaussienne standard, $H=0.8$

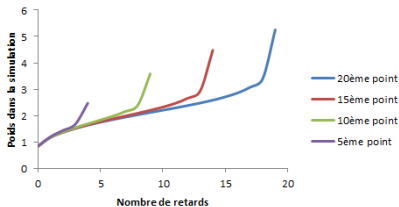


FIGURE – $\sigma = 1$. Comprendre, pour $H = 0.8$, le 20ème point simulé attribue un poids de 0.83 au 20ème tirage de v.a. gaussienne standard, puis 1.17 au 19ème, etc. jusqu'à 5.25 au premier : il s'agit d'un cas de mémoire longue.

Mouvement brownien fractionnaire - Simulation

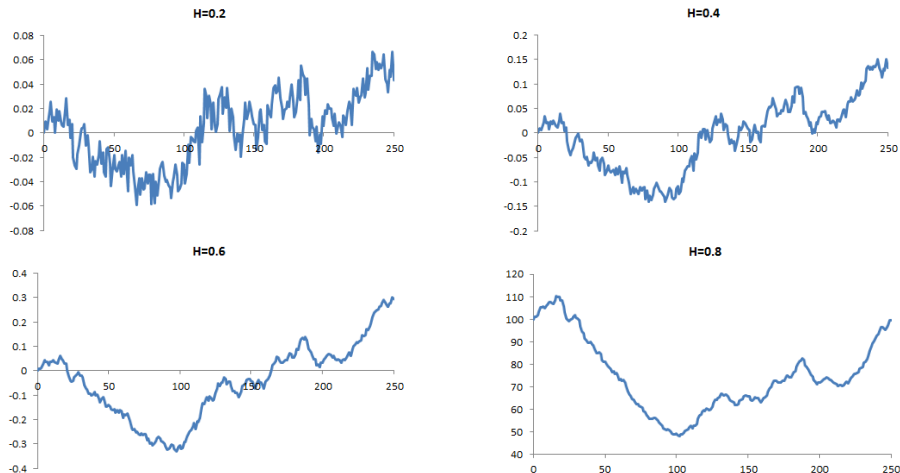


FIGURE – MBf. σ correspond à une vol annuelle de 20%. La mémoire longue, lorsque H est grand, favorise une plus grande amplitude du signal (mais avec moins d'oscillations).

Mouvement brownien fractionnaire - Simulation

Pour une simulation de prix d'action par exemple, on utilise d'habitude le mouvement brownien géométrique :

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_{1/2}(t),$$

dont la solution s'obtient en appliquant la formule d'Itô

$$d(f(X_t, t)) = \frac{\partial f}{\partial t}(X_t, t)dt + \frac{\partial f}{\partial x}(X_t, t)dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X_t, t)\sigma_t^2 dt$$

à $f(S_t) = \ln(S_t)$:

$$S_t = S_0 \exp \left(\mu t + \sigma B_{1/2}(t) - \frac{\sigma^2}{2} t \right).$$

Mouvement brownien fractionnaire - Simulation

Dans le cas général, c'est-à-dire fractal, la formule d'Itô devient :

$$d(f(X_t, t)) = \frac{\partial f}{\partial t}(X_t, t)dt + \frac{\partial f}{\partial x}(X_t, t)dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(X_t, t)\sigma_t^2(dt)^{2H}$$

et l'équation du MBf géométrique

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_H(t)$$

a pour solution

$$S_t = S_0 \exp \left(\mu t + \sigma B_H(t) - \frac{\sigma^2}{2} t^{2H} \right).$$

Attention !

Si vous voulez estimer un MBf géométrique, il faut bien estimer H sur le logarithme népérien du prix et non le prix directement.

Mouvement brownien fractionnaire - Simulation

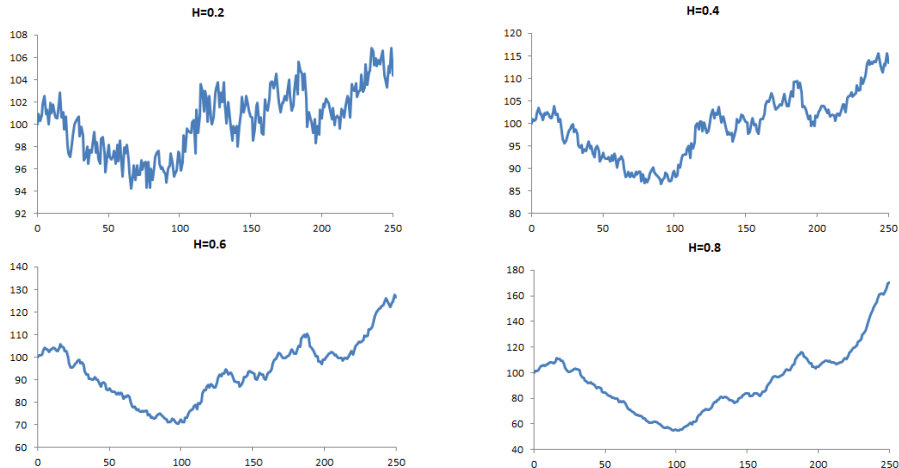


FIGURE – Mouvement brownien fractionnaire géométrique (si $H = 1/2$, mouvement brownien géométrique utilisé en finance) avec même suite de variables pseudo-aléatoires. σ correspond à une vol annuelle de 20%.

Mouvement brownien fractionnaire

Quelle utilité ?

- Généralise mouvement brownien et étend le calcul stochastique (applications au pricing d'options par exemple : d'où la nécessité de savoir simuler un MBf).
- Une dynamique plus riche que le mouvement brownien : mémoire longue, singularités.

Limites du MBf en finance

- On a vu que l'exposant de Hurst diminue les années précédant une singularité prononcée (1987, 2008), contrairement à la volatilité qui n'anticipe pas. Mais le MBf suppose une fractale homogène, i.e. avec des singularités partout de même exposant de Hölder égal à l'exposant de Hurst. Il faudrait donc plutôt une distribution d'exposants plutôt qu'une seule valeur (solution : processus localement autosimilaire : mouvement brownien multifractionnaire).
- Estimation sensible au pas de temps : la dynamique des In-prix ne doit pas être un MBf.
- Processus non stationnaire (solution : transformée de Lamperti).
- Rendements gaussiens (solution : mouvement Lévy-stable fractionnaire).

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire
 - Le mouvement brownien fractionnaire
 - Calibration de Black-Scholes fractionnaire
 - Au delà du mbf

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Calibration Black-Scholes fractionnaire

Dans la littérature, la formule de Black et Scholes dans le cadre du MBF devient :

$$C(S_0, K, r, T, \sigma) = S_0 \mathcal{N}(d_1) - Ke^{-rT} \mathcal{N}(d_2),$$

avec $d_1 = \frac{1}{\sigma T^H} \left[\ln \left(\frac{S_0}{K} \right) + rT + \frac{1}{2} \sigma^2 T^{2H} \right]$ et $d_2 = d_1 - \sigma T^H$.

Ce prix est conditionnel au prix en $t = 0$ mais ne tient pas compte du passé $t < 0$. Pourtant, si H élevé et passé haussier, avenir est plus probablement haussier que baissier. Difficulté car alors prix ne seraient pas martingales (contradiction avec hypothèse des marchés efficients levée si on a vol stochastique fractionnaire mais pas processus des prix, par exemple, mais est-ce nécessaire de lever cette contradiction ?).

Avantage de ce modèle : Dans l'approche du brownien standard, on doit avoir la même volatilité annuelle à toutes les maturités, ie une maturité mise à l'échelle de $\sigma_{BS} T^{1/2}$; ici la volatilité mise à l'échelle est de la forme $\sigma_{fBM} T^H$, ce qui, dans le cadre d'une annualisation standard donne $\sigma_{fBM} T^{H-\frac{1}{2}}$, soit une fonction puissance et pas seulement une term-structure constante, ce qui correspond aux observations empiriques sur les options.

Calibration Black-Scholes fractionnaire

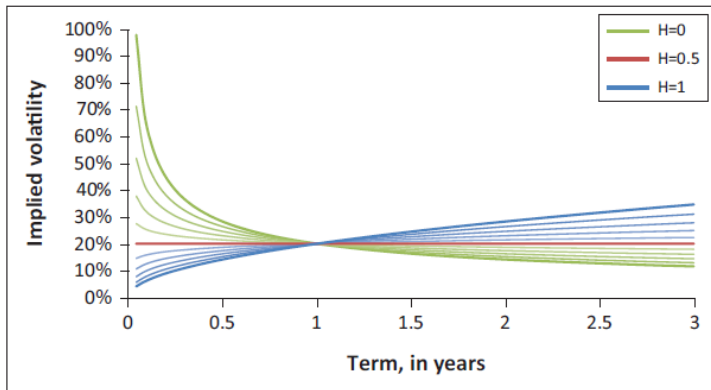


FIGURE 1: Possible implied volatility term structures in a fractional Black-Scholes market for different Hurst exponents and $\sigma_f = 20\%$.

FIGURE — Source : Flint, Maré.

Calibration Black-Scholes fractionnaire

Deux paramètres à calibrer : σ et H .

- Possibilité de fixer l'un ou l'autre.
- Possibilité d'optimiser conjointement en ajoutant pénalité sur les variation de l'un et/ou l'autre :

$$\min_{\sigma(K_i, T_j), H(K_i, T_j)} \sum_{i,j} [\sigma(K_{i+1}, T_j) - \sigma(K_i, T_j)]^2 + \lambda [H(K_{i+1}, T_j) - H(K_i, T_j)]^2 \\ + \mu [\sigma(K_i, T_{j+1}) - \sigma(K_i, T_j)]^2 + \nu [H(K_i, T_{j+1}) - H(K_i, T_j)]^2$$

sous contrainte $\forall i, j, C(\sigma(K_i, T_j), H(K_i, T_j), K_i, T_j) = C^{\text{marché}}(K_i, T_j)$. En pratique, pour prendre en compte la contrainte, il suffit de remplacer les $H(K_i, T_j)$ par une fonction $\mathcal{H}_{i,j}$ de $\sigma(K_i, T_j)$ telle que $C(\sigma(K_i, T_j), \mathcal{H}_{i,j}(\sigma(K_i, T_j)), K_i, T_j) = C^{\text{marché}}(K_i, T_j)$, ce qui divise par deux la taille de l'espace sur lequel on minimise la fonction objectif : on détermine, par une fonction d'optimisation, la fonction $\mathcal{H}_{i,j}$, puis on résout :

$$\min_{\sigma(K_i, T_j)} \sum_{i,j} [\sigma(K_{i+1}, T_j) - \sigma(K_i, T_j)]^2 + \lambda [\mathcal{H}_{i+1,j}(\sigma(K_{i+1}, T_j)) - \mathcal{H}_{i,j}(\sigma(K_i, T_j))]^2 \\ + \mu [\sigma(K_i, T_{j+1}) - \sigma(K_i, T_j)]^2 + \nu [\mathcal{H}_{i,j+1}(\sigma(K_i, T_{j+1})) - \mathcal{H}_{i,j}(\sigma(K_i, T_j))]^2.$$

• ...

Calibration Black-Scholes fractionnaire

- ...
- Possibilité de fixer un modèle paramétrique. En effet, en général, pour les options sur actions très dans la monnaie, on observe que nappes s'aplatissent pour maturités grandes, ce qui correspond à $H < 0.5$, alors que pour grands strikes ou options sur devises, $H > 0.5$ est acceptable. Donc H devrait dépendre du strike :
 - Approche FBSI (Fractional Black-Scholes inspired) : variance selon strike ($v(K)$) suit modèle SVI, nappes de variance s'écrivent $v(K)T^{2H(K)}$ avec $H(K) = \beta_0 + \beta_1 K + \beta_2 K^2$.
 - Approche concurrente LC (Li et Chen), pas vraie calibration : pour un strike donné, on régresse le log de la vol BS à travers les maturités sur le log de la vol fractionnaire $\sigma_{fBm} T^{H-\frac{1}{2}}$.

Calibration Black-Scholes fractionnaire

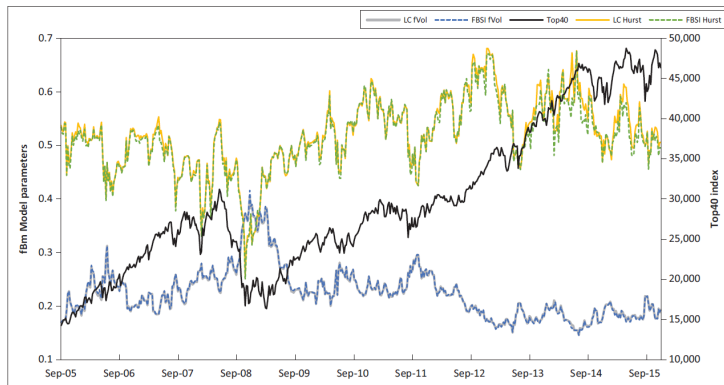


FIGURE 4: Top40 index performance plotted with the ATM fractional volatility and Hurst exponents from the calibrated FBSI and LC models, respectively, September 2005 to November 2015.

FIGURE – Plus grande stabilité avec approche mbf qu'avec BS standard (Top40 est un indice actions Sud-Africain). Source : Flint, Maré.

Calibration Black-Scholes fractionnaire

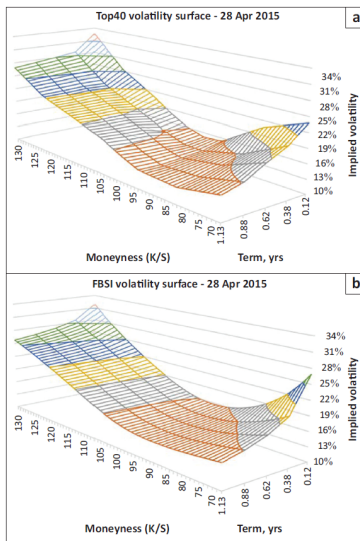


FIGURE — Source : Flint, Maré.

Calibration Black-Scholes fractionnaire

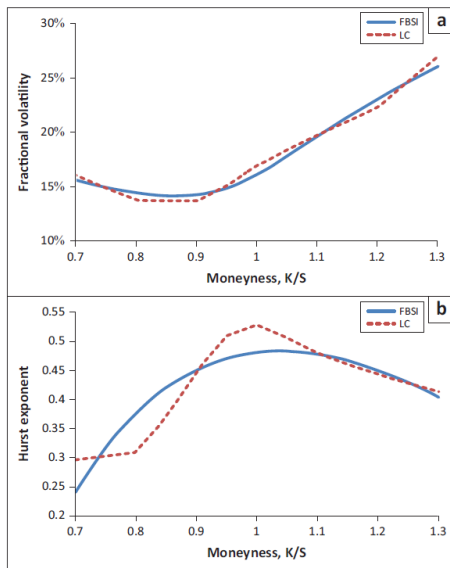


FIGURE 9: Calibrated (a) fractional volatility and (b) Hurst exponent skews from the FBSI and LC models as at 30 May 2011.

1 Algorithmes d'optimisation

2 Calibration de volatilité implicite

- Modèle de Black-Scholes, volatilité implicite, volatilité locale
- Modèles à volatilité stochastique
- Densité risque-neutre
- Paramètres implicites d'un mouvement Lévy-stable fractionnaire
 - Le mouvement brownien fractionnaire
 - Calibration de Black-Scholes fractionnaire
 - Au delà du mbf

3 Théorie de la calibration

4 L'approche statistique

5 Exemples de calibrations

Mouvement brownien multifractionnaire

Si l'exposant de Hurst dépend du temps selon une fonction höldérienne d'exposant strictement positif, $H : \mathbb{R}^+ \rightarrow (0, 1)$, alors on définit un mouvement brownien multifractionnaire $B_{H(t)}(t)$, qui généralise le MBf. La covariance est :

$$\begin{cases} \mathbb{E}\{B_H(t)B_H(s)\} = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) & \text{si MBf} \\ \mathbb{E}\{B_{H(t)}(t)B_{H(s)}(s)\} = \frac{\sigma^2}{2}(|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)}) & \text{si MBm.} \end{cases}$$

Pour son estimation, on peut :

- conserver l'estimation par fenêtres glissantes comme précédemment ;
- estimer une approximation de la densité beaucoup plus précise (locale et non moyenne sur petite fenêtre) grâce aux ondelettes, comme on le verra plus tard.

Mouvement brownien fractionnaire - Lampertisation

Définition

Un processus $\{Y_t, t \in \mathbb{R}^+\}$ est dit **stationnaire** si $\{(S_\tau Y)_t, t \in \mathbb{R}^+\}$ et $\{Y_t, t \in \mathbb{R}^+\}$ ont même distribution pour tout $\tau \in \mathbb{R}$, où $(S_\tau Y)_t = Y_{t+\tau}$ est l'opérateur de translation.

Pourquoi préfère-t-on des processus stationnaires ?

- Une part des théorèmes statistiques nécessite cette condition (théorie des valeurs extrêmes, ...).
- Correspond à une réalité en finance : le mouvement brownien géométrique a du sens pour une action, mais pas pour une obligation dont on sait qu'elle doit rembourser 100 à maturité (ou pour un taux) et qui favorise plutôt des oscillations autour d'une moyenne.
- Certaines opérations sont plus facile avec un processus ou avec son dual par transformée de Lamperti.

Mouvement brownien fractionnaire - Lampertisation

Exemple

Le processus d'Ornstein-Uhlenbeck, r_t , défini par l'équation :

$$dr_t = -\theta(r_t - \mu)dt + \sigma dW_t,$$

dont la solution est (si $r_0 = 0$) :

$$r_t = \mu + \frac{\sigma}{\sqrt{2\theta}} W_{\exp(2\theta t)} e^{-\theta t}.$$

Pour un taux, ce modèle s'appelle Vasicek, avec :

- θ la vitesse de retournement ;
- μ le taux moyen de long terme ;
- σ la volatilité.

L'amplitude du signal diminue puis augmente à mesure que θ augmente.

Mouvement brownien fractionnaire - Lampertisation

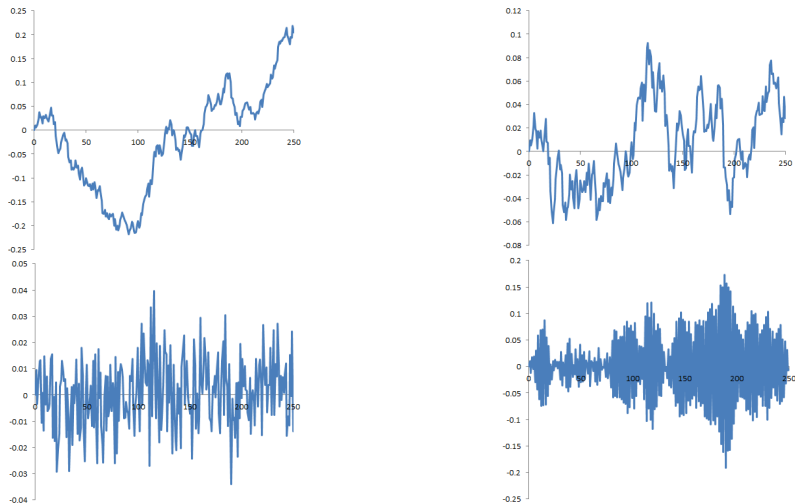


FIGURE – OU avec $\sigma = 20\%/\sqrt{256}$, θ égal (de gauche à droite puis haut en bas) à 0 (MB), 0.1, 1 et 2, $\mu = 0$, même suite de nombres pseudo-aléatoires.

Mouvement brownien fractionnaire - Lampertisation

Définition

- On note $\mathcal{D}_{H,\lambda}$ l'opérateur de dilatation, pour $H > 0$ et $\lambda > 0$:

$$(\mathcal{D}_{H,\lambda}X)_t = \lambda^{-H}X_{\lambda t}$$

- Conformément à ce que l'on a vu pour les fractales, un processus X est H -autosimilaire si X_t et $(\mathcal{D}_{H,\lambda}X)_t$ ont même distribution pour tout $\lambda > 0$.

Proposition

Si X est H -autosimilaire, alors $t^{H'}X_t$ est $H + H'$ -autosimilaire.

Mouvement brownien fractionnaire - Lampertisation

Définition

Pour $H > 0$, la transformée de Lamperti, \mathcal{L}_H , est un opérateur transformant un processus $\{Y_t, t \in \mathbb{R}^+\}$ selon la règle suivante :

$$(\mathcal{L}_H Y)_t = t^H Y_{\ln t}.$$

La transformation inverse, \mathcal{L}_H^{-1} , du processus $\{X_t, t \in \mathbb{R}^+\}$ est définie par :

$$(\mathcal{L}_H^{-1} X)_t = e^{-Ht} X_{\exp(t)}.$$

Proposition

- La lampertisation est inversible : $(\mathcal{L}_H^{-1} \mathcal{L}_H Y)_t = (\mathcal{L}_H \mathcal{L}_H^{-1} Y)_t = Y_t$.
- On peut montrer que $(\mathcal{L}_{H_2}^{-1} \mathcal{L}_{H_1} Y)_t = e^{-(H_2 - H_1)t} Y_t$.
- De même $(\mathcal{L}_{H_2} \mathcal{L}_{H_1}^{-1} X)_t = t^{H_2 - H_1} X_t$.

Mouvement brownien fractionnaire - Lampertisation

En fait, stationnarité et autosimilarité sont liées grâce à la lampertisation.

Théorème (Théorème de Lamperti)

Si Y est un processus stationnaire, alors $\mathcal{L}_H Y$ est H -autosimilaire.

Réciproquement, si X est H -autosimilaire, alors $\mathcal{L}_H^{-1} X$ est stationnaire.

- 1 En effet, montrons déjà que pour tout $\lambda > 0$: $\mathcal{L}_H^{-1} \mathcal{D}_{H,\lambda} \mathcal{L}_H = \mathcal{S}_{\ln \lambda}$.
- 2 Ensuite, si Y est stationnaire, alors, pour tout $\lambda > 0$, Y est distribué comme $\mathcal{S}_{\ln \lambda} Y$, c'est-à-dire $\mathcal{L}_H^{-1} \mathcal{D}_{H,\lambda} \mathcal{L}_H Y$. Donc, en posant $X = \mathcal{L}_H Y$, on a même distribution pour X et pour $\mathcal{D}_{H,\lambda} X$, et ce quel que soit $\lambda > 0$. Donc X est autosimilaire.
- 3 La réciproque est prouvée par un raisonnement symétrique.

Mouvement brownien fractionnaire - Lampertisation

Alors, les propriétés des processus autosimilaires peuvent être obtenues des propriétés de la transformée stationnaire. De là à appliquer la théorie des valeurs extrêmes (ou autre) au MBf, il n'y a qu'un pas.

Proposition

Soit $R_X(t, s) = \mathbb{E}\{X_t X_s\}$.

- Alors $R_{\mathcal{L}_H^{-1}X}(t, s) = e^{-H(t+s)} R_X(e^t, e^s)$.
- De même $R_{\mathcal{L}_HY}(t, s) = (ts)^H R_Y(\ln(t), \ln(s))$.
- Donc, si Y est stationnaire, il existe une fonction γ_Y telle que $R_Y(t, s) = \gamma(t - s)$ et alors $R_{\mathcal{L}_HY}(t, s) = (ts)^H \gamma_Y(\ln(t/s))$.
- Par le théorème de Lamperti, si X est H -autosimilaire, X a nécessairement une fonction de covariance de la forme $R_X(t, s) = (ts)^H c_H(t/s)$, où c_H est une fonction liée à la transformée de Lamperti de X .

Mouvement brownien fractionnaire - Lampertisation

Exemple

- Le processus d'Ornstein-Uhlenbeck est la transformée de Lamperti d'un mouvement brownien. Le MB est $1/2$ -autosimilaire, l'OU est stationnaire.
- Le MBF permet de généraliser l'OU, avec un OU fractionnaire, par lampertisation.

Mouvement brownien fractionnaire - Lampertisation

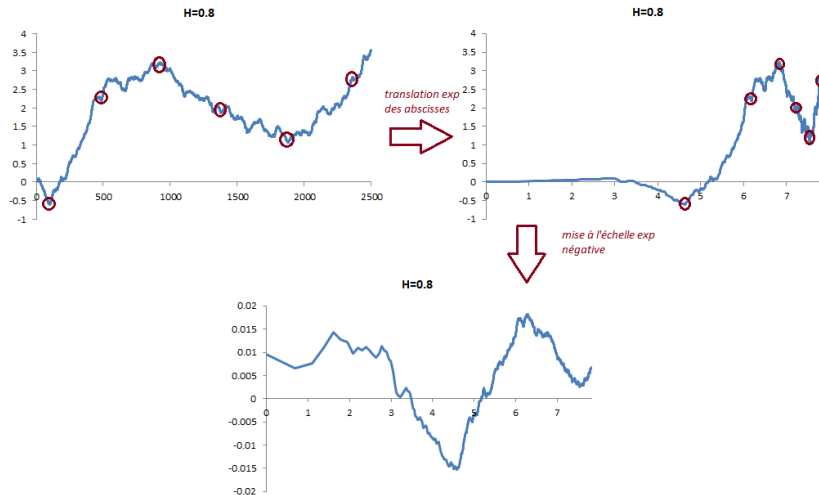


FIGURE – OU généralisé : lampertisation d'un MBf avec $\sigma = 20\%/\sqrt{256}$, $H = 0.8$.

Mouvement brownien fractionnaire - Lampertisation

Si X est un MBf d'exposant de Hurst H , alors

$$R_X(t, s) = \frac{\sigma^2}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

On peut l'écrire sous la forme $R_X(t, s) = (ts)^H c_H(t/s)$ avec

$$c_H(\lambda) = \frac{\sigma^2}{2} [\lambda^H + \lambda^{-H} (1 - |1 - \lambda|^{2H})].$$

La transformée inverse de Lamperti de X vérifie alors

$$R_{\mathcal{L}_H^{-1}X}(t, s) = e^{-H(t+s)} \frac{\sigma^2}{2} (e^{2Ht} + e^{2Hs} - |e^t - e^s|^{2H}).$$

Alors, $\mathcal{L}_H^{-1}X$ étant stationnaire, sa fonction de covariance stationnaire est

$$\gamma_{\mathcal{L}_H^{-1}X}(\tau) = R_{\mathcal{L}_H^{-1}X}(\cdot, \cdot + \tau) = \sigma^2 \left(\cosh(H|\tau|) - \left[\sinh\left(\frac{|\tau|}{2}\right) \right]^{2H} \right).$$

Pour $H = 1/2$, on reconnaît la fonction de covariance stationnaire d'un OU : $\sigma^2 e^{-|\tau|/2}$.

Mouvement brownien fractionnaire - Lampertisation

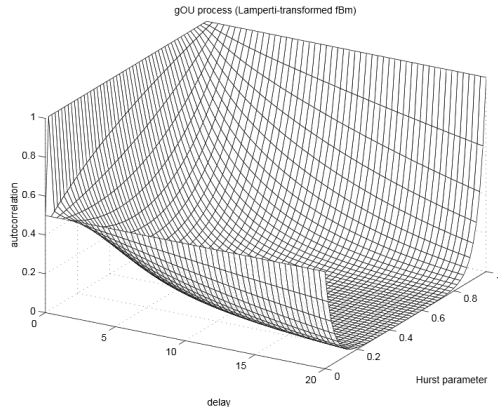


Figure 3: Stationary covariance function of generalized Ornstein-Uhlenbeck processes — “Delampertizing” fractional Brownian motion (fBm) ends up with a stationary process, referred to as a *generalized Ornstein-Uhlenbeck* (gOU) process, whose covariance function is plotted here as a function of the Hurst parameter H . It is worth noting that this covariance decays exponentially fast for any $H \in (0, 1)$, indicating that gOU is always *short-range dependent*.

FIGURE – Source : Flandrin, Borgnat, Amblard.

Mouvement brownien fractionnaire - Lampertisation

Simulation

Pour simuler un OU **fractionnaire** (resp. généralisé), on aura une approche directe (resp. une lampertisation d'un MBf mais pas déjà simulé car on a sinon de grands trous dans la discrétisation). Pour simuler directement l'OUf, on simule les incréments :

$$dr_t = -\theta(r_t - \mu)dt + \sigma dB_H(t).$$

Mouvement brownien fractionnaire - Lampertisation

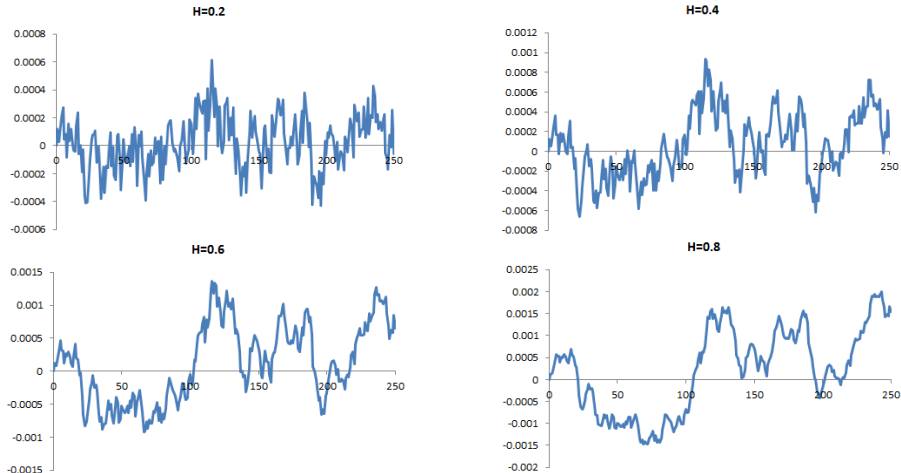


FIGURE – OU fractionnaire pour plusieurs H avec $\sigma = 20\%/\sqrt{256}$, $\theta = 0.1$, $\mu = 0$, même suite de nombres pseudo-aléatoires que pour la simulation des MBf précédemment montrée.

Mouvement brownien fractionnaire - Lampertisation

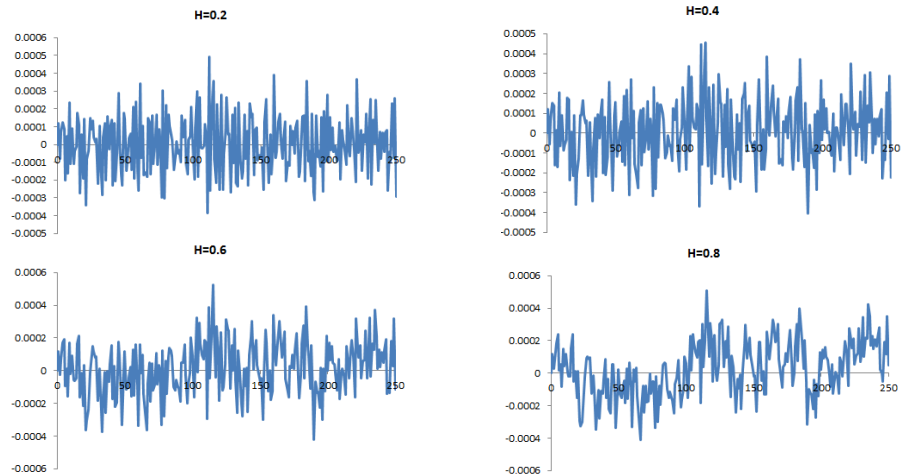


FIGURE – OU fractionnaire pour plusieurs H avec $\sigma = 20\%/\sqrt{256}$, $\theta = 1$, $\mu = 0$, même suite de nombres pseudo-aléatoires. Ressemblance cas ($\theta = 1$, $H = 0.8$) et ($\theta = 0.1$, $H = 0.2$).

Mouvement Lévy-stable fractionnaire

Loi d'échelle donnée par H . En fait, deux manières (non exclusives) de modifier loi d'échelle d'un brownien standard :

- en ajoutant de la dépendance (approche mBf),
- en transformant loi gaussienne en loi stable (approche Lévy-stable).

Définition

Une loi de probabilité F est **stable** signifie que pour toutes variables aléatoire X , Y et Z de fonction de répartition F et pour tout $\alpha, \beta \geq 0$, il existe γ et δ tels que :

$$\alpha X + \beta Y =_{\text{loi}} \gamma Z + \delta.$$

Mouvement Lévy-stable fractionnaire - Loi stable

Théorème

Théorème central limite Si l'espérance et la variance de la variable aléatoire X_1 sont finies, alors, on a la loi des écarts à la moyenne de S_n :

$$\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n\text{Var}(X_1)}} \xrightarrow{\text{loi}, n \rightarrow \infty} \mathcal{N}(0, 1),$$

où $\mathcal{N}(\mu, \sigma^2)$ est la loi normale d'espérance μ et de variance σ^2 .

Théorème

Soit a_n et $b_n > 0$ les termes d'une suite telle que :

$$\frac{S_n - a_n}{b_n} \xrightarrow{\text{loi}, n \rightarrow \infty} G,$$

où G est une fonction de répartition non-dégénérée.

Alors G est du même type qu'une loi stable.

Mouvement Lévy-stable fractionnaire - Loi stable

Proposition

La fonction caractéristique d'une loi stable de paramètres $\alpha, \beta, \mu, \gamma$ (avec $\gamma > 0$) est nécessairement de la forme suivante :

- pour une loi 2-stable (cas particulier du cas suivant) on trouve une loi gaussienne, avec γ la moitié de la variance :

$$\phi_X(t) = \exp(-\gamma t^2) ;$$

- pour $\alpha \in [0, 2]$ mais non-entier et $\beta \in [-1, 1]$:

$$\phi_X(t) = \exp \left(i\mu t - \gamma |t|^\alpha \left[1 - i\beta \operatorname{signe}(t) \tan \left[\frac{\pi\alpha}{2} \right] \right] \right) ;$$

- pour $\alpha = 1$ et $\beta \in [-1, 1]$, on retrouve une loi de Cauchy :

$$\phi_X(t) = \exp \left(i\mu t - \gamma |t| \left[1 + i\beta \operatorname{signe}(t) \frac{2}{\pi} \log |t| \right] \right).$$

Mouvement Lévy-stable fractionnaire

- The fLsm is defined by the integral

$$W_t^{\alpha,m} = \int_{-\infty}^{\infty} ((t-s)_+^m - (-s)_+^m) dW_s^{\alpha,0},$$

where $W_s^{\alpha,0}$ is a symmetric Lévy α -stable motion, that is to say a stochastic process with i.i.d. increments of symmetric Lévy α -stable distribution (Weron *et al.*).

- The characteristic function of a centred symmetric stable distribution ($u \mapsto \exp(-\gamma|u|^\alpha)$) helps to estimate α , and to define a process.⁷
- The Hurst exponent can be decomposed in a memory and a distribution components :

$$H = m + \frac{1}{\alpha}.$$

Standard Brownian motion has $m = 0$ (no memory) and $\alpha = 2$.

7. The α -stable distribution is particularly relevant since the sum of two independent α -stable variables of scale parameter σ is an α -stable variable of scale parameter $2^{1/\alpha}\sigma$. Therefore, in this i.i.d. setting, $1/\alpha$ plays a role of scaling parameter in the same manner as does the Hurst exponent in the fBm.

Mouvement Lévy-stable fractionnaire - Estimation

If fLsm, use characteristic function for a well-chosen u , like $\pi/4Q_{90\%}$, where Q is the quantile function of $|\hat{Z}_t - \hat{Z}_{t-1}|$.

$$\hat{\alpha} = \frac{1}{\ln(2)} \ln \left(\frac{\ln \left(\frac{1}{T} \sum_{t=1}^T \cos(2u(Z_t - Z_{t-1})) \right)}{\ln \left(\frac{1}{T} \sum_{t=1}^T \cos(u(Z_t - Z_{t-1})) \right)} \right),$$

$$\hat{\sigma} = -|u|^{-\hat{\alpha}} \ln \left(\frac{1}{T} \sum_{t=1}^T \cos(u(Z_t - Z_{t-1})) \right),$$

$$\hat{H} = \frac{1}{\ln(2)} \ln \left(\frac{\frac{2}{T} \sum_{t=1}^{\lfloor T/2 \rfloor} |Z_{2t} - Z_{2(t-1)}|}{\frac{1}{T} \sum_{t=1}^T |Z_t - Z_{t-1}|} \right).$$

Mouvement Lévy-stable fractionnaire - Calibration

- Décomposition de H en $m + 1/\alpha$ change un peu prix des calls (distribution non gaussienne) non conditionnels (mais change aussi beaucoup prix conditionnels car la mémoire change!).
- On peut sinon imposer une structure (à travers maturités et strikes) non plus seulement pour H mais aussi pour m et pour α , ce qui peut fournir des modèles plus riches, plus en adéquation avec réalité.
- Occasion d'enfin se passer d'une loi gaussienne : tout est résumé dans le α qui donne une structure plus générale que la distribution gaussienne (qui en est un cas particulier). On pourrait aussi à peu de frais ajouter asymétrie.
- Autre manière de faire pour analyser le marché (à mi-chemin entre calibration et estimation) : quel α constant permet d'avoir vol implicite la plus plate possible ? Sorte d' α moyen du marché (très faible : anticipation de crises ; donc risque extrême implicite).

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration**
 - Régularisation
 - Estimation ou calibration ?
- 4 L'approche statistique
- 5 Exemples de calibrations

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration**
 - Régularisation
 - Problèmes inverses mal posés
 - Régularisation de Tikhonov
 - Estimation ou calibration ?
- 4 L'approche statistique
- 5 Exemples de calibrations

Problèmes inverses mal posés

Problème direct

Détermine effet à partir des causes. Exemples :

- Prix d'une option déterminé à partir de prix initial et paramètres d'une dynamique (volatilité).
- Détermination du temps pris par un rayon X (dans une direction donnée) pour traverser un corps dont on connaît la composition exacte. Pour θ, z , on calcule $t(\theta, z) = \int f(r, \theta, z) dr$.

Problème inverse

Induit les causes à partir d'effets. Exemples :

- Volatilité implicite.
- Tomographie : à partir des temps pris par un rayon X (dans plein de directions) pour traverser un corps, on induit sa géographie. A partir des $t(\theta, z) = \int f(r, \theta, z) dr$, comment induire $f(r, \theta, z)$?

Plus compliqué que problème direct car nécessite réciproque de fonction.

Problèmes inverses mal posés

Définition

Un problème $Tx = y$, où T est un opérateur, pas forcément linéaire, est bien posé si :

- pour y observé, x existe et est unique ;
- x dépend continûment de y .

Exemples :

- T fonction de valorisation de Black-Scholes, x vecteur contenant la volatilité du modèle et le prix initial de l'action, y prix de marché de l'option.
- Problème inverse mal posé de calcul de dérivée (voir ci-après) : $\int_{-\infty}^u x(t)dt = y(u)$, où $T = \int_{-\infty}^u .dt$.

Ce que va illustrer l'exemple de la dérivée

- Dans problème mal posé, les erreurs des observations sont amplifiées dans la solution.
- L'information *a priori* sur la solution (ici f''') et le niveau d'erreur, permet d'optimiser la stabilité de la solution (sans recours à régularisation).

Problèmes inverses mal posés - Exemple de la dérivation

- On cherche dérivée $f'(x)$, dérivée de $f(x)$, calculée à partir de fonction bruitée $f^\delta(x)$ qui est telle que $|f(x) - f^\delta(x)| < \delta$ pour tout x .
- f est dérivable, mais pas forcément f^δ . Ainsi, f' déterminé à partir de f va dépendre de manière non continue de x : mal posé.
- Approximation : $f'(x) \approx \frac{f^\delta(x+h) - f^\delta(x-h)}{2h}$, pour un h arbitrairement petit.
 - Si h trop grand, erreur de discrétisation. Formule de Taylor-Lagrange : pour $\eta \in \{-1, 1\}$, $\left| f(x + \eta h) - \left[f(x) + \eta h f'(x) + \frac{h^2}{2} f''(x) \right] \right| \leq \frac{h^3}{6} C$, où C est un majorant de $|f'''|$. En appliquant la formule en $\eta = 1$ et en ôtant celle en $\eta = -1$, on obtient :

$$\left| f'(x) - \frac{f(x+h) - f(x-h)}{2h} \right| \leq h^2 C / 6.$$

- Si h est trop petit, les erreurs de données sont amplifiées :

$$\left| \frac{f^\delta(x+h) - f^\delta(x-h)}{2h} - \frac{f(x+h) - f(x-h)}{2h} \right| \approx \delta/h.$$

Ainsi, erreur totale est de l'ordre de $\delta/h + h^2 B$, où B est une constante (ne dépend pas de h ni de δ). Cette erreur est minimale pour $h = (\delta/2B)^{1/3}$, avec une erreur alors de l'ordre de $\delta^{2/3}$.

Problèmes inverses mal posés - Exemple de la dérivation

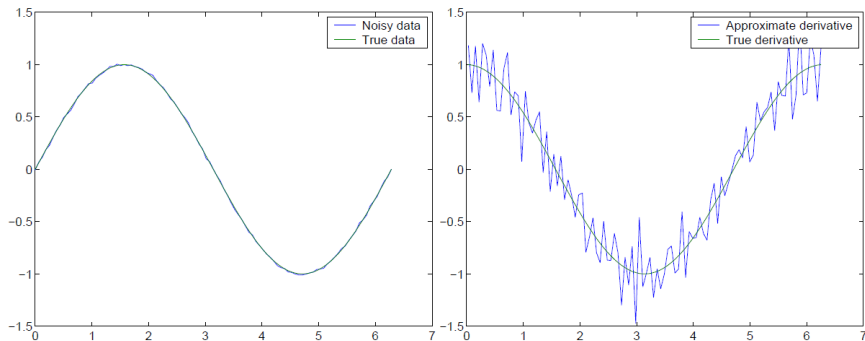


Figure 6.1: Gauche: les données exactes et bruitées. Droite: la dérivée exacte de f et la dérivée approchée de f^δ avec $h = 0.03$.

FIGURE – Source : Tankov.

Problèmes inverses mal posés - Exemple de la dérivation

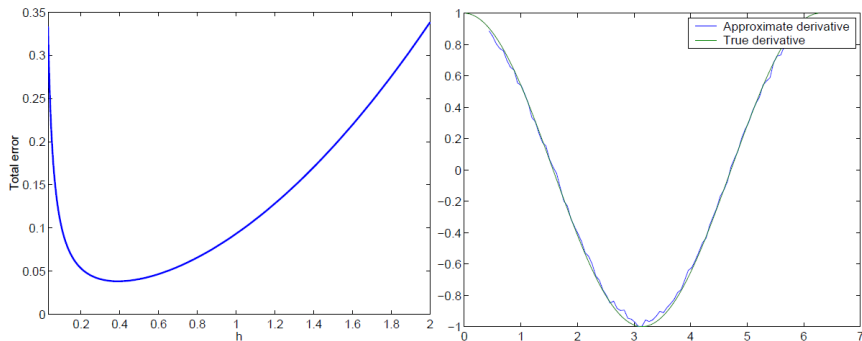


Figure 6.2: Gauche: Erreur totale (données + discrétisation) en fonction de pas de discrétisation h . Droite: la dérivée exacte de f et la dérivée approchée de f^δ avec $h_{opt} = 0.39$.

FIGURE – Source : Tankov.

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration**
 - Régularisation
 - Problèmes inverses mal posés
 - Régularisation de Tikhonov
 - Estimation ou calibration ?
- 4 L'approche statistique
- 5 Exemples de calibrations

Le problème $Tx = y$, où T est un opérateur linéaire (possible d'adapter le cadre à T non-linéaire) a pour solution, si elle existe,

$$T^\bullet y = (T^* T)^{-1} T^* y,$$

où T^* est l'adjoint de T , ie le conjugué de sa transposée.

Si problème mal posé (qui occasionne d'avoir x tendant vers l'infini pour certains points), notamment à cause de bruit (on observe y^δ au lieu de y , avec $\|y^\delta - y\| \leq \delta$) on régularise, ie on résout le problème suivant :

$$\operatorname{arginf}_x \left\{ \|Tx - y^\delta\|^2 + \alpha \|x\|^2 \right\},$$

où :

- α dépend positivement de δ : si données peu fiables, régularisation doit être forte,
- Dès que $\alpha > 0$, il y a une perte de précision par rapport aux moindres carrés mais le problème devient bien posé,
- α est choisi de telle sorte que quand $\delta \rightarrow 0$, la solution converge vers $T^\bullet y$.

Problème d'optimisation, condition au premier ordre donne la solution :

$$x_\alpha^\delta = (T^* T + \alpha I)^{-1} T^* y^\delta.$$

Choix de α – Méthode *a priori*

Théorème

Si $\alpha(\delta)$ est choisi tel que $\lim_{\delta \rightarrow 0} \alpha(\delta) = 0$ et $\lim_{\delta \rightarrow 0} \frac{\delta^2}{\alpha(\delta)} = 0$, alors

$$\lim_{\delta \rightarrow 0} \|x_{\alpha(\delta)}^{\delta} - T^{\bullet}y\| = 0.$$

De plus, s'il existe z tel que $T^{\bullet}y = T^*Tz$ alors la convergence asymptotique maximale est atteinte pour $\alpha(\delta) = C\delta^{2/3}$ et alors

$$\|x_{\alpha}^{\delta} - T^{\bullet}y\| = \mathcal{O}(\delta^{2/3}).$$

En fait, on peut décomposer l'erreur globale en variance et biais :

$$\|x_{\alpha}^{\delta} - T^{\bullet}y\| \leq \|x_{\alpha}^{\delta} - x_{\alpha}\| + \|x_{\alpha} - T^{\bullet}y\|.$$

La variance décroît en α et le biais croît en α . Il faut trouver un bon compromis.

Choix de α – Méthode *a posteriori*

La fonction de discrédance donne l'écart entre les données et ce que l'on retrouverait en utilisant la solution (par exemple écart entre prix de marché et prix de modèle) :

$$\varepsilon(\alpha) = \|Tx_{\alpha}^{\delta} - y^{\delta}\|.$$

On choisit alors le α maximal tel que la discrédance est acceptable, par exemple légèrement supérieure à l'erreur de données. Cela est possible car ε est croissant, continue, avec $\lim_{\alpha \rightarrow 0} \varepsilon(\alpha) \leq \delta$ et $\lim_{\alpha \rightarrow \infty} \varepsilon(\alpha) = \|y^{\delta}\|$.

Comment déterminer δ

- Si on suppose x_{α}^{δ} suffisamment proche de x , la discrédance maximale nous donne une estimation de δ (mais dans ce cas δ dépend de x_{α}^{δ} qui dépend de δ ...). Algorithme itératif.
- Supposons un bruit gaussien, donc non borné, mais on veut avoir son amplitude moyenne (écart-type) σ . La discrédance médiane multipliée par 1.293 est alors un estimateur robuste de σ .

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration**
 - Régularisation
 - Estimation ou calibration ?
- 4 L'approche statistique
- 5 Exemples de calibrations

Estimation ou calibration ?

• Calibration :

- problème inverse (le problème direct est facile : c'est le pricing) mal posé (inverse très instable dans certains cas, ie T^{-1} non continu, par exemple si pricing avec modèle fBm conditionnel : pour paramètre $H = 1/2$, on ne prend pas en compte le passé, contrairement à $H > 1/2$).
- S'écrit sous la forme $Tx = y$, avec T opérateur non-linéaire (fonction de pricing), x vecteur de paramètres et y vecteur de prix de marché d'options.
- Régularisation permet de réduire erreurs non pas des données mais dues à l'instabilité de l'inversion de T . Pas utile si inverse stable (simple Black-Scholes).
- Données y sont jugées vraies (erreur bid-ask près) et convergence s'entend donc par le fait que l'écart de y à réalité tend vers 0 (grande liquidité).

• Estimation :

- Modèle aussi de la forme $Tx = y$, mais ici c'est la donnée y qui est jugée peu fiable.
- On régularise pour éviter overfitting (et ainsi avoir une qualité prédictive accrue).
- Convergence pour un nombre infini d'observations.

La différence pourrait se résumer à déterministe (calibration) vs aléatoire (estimation), mais il y a aussi de l'aléatoire dans les modèles de valo.

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 L'approche statistique**
 - Régularisation et statistiques non-paramétriques
- 5 Exemples de calibrations

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 **L'approche statistique**
 - Régularisation et statistiques non-paramétriques
 - Surapprentissage et sélection de modèle
 - Méthodes de lissage
 - LASSO
- 5 Exemples de calibrations

Overfitting

Définition

Surapprentissage : *Ajout de variables explicatives (ou paramètres) peu pertinentes : améliore précision in-sample mais la détériore out-of-sample.*

C'est le danger en statistiques non-paramétriques ou paramétriques avec grand nombre de coefficients. Corrigé avec :

- régularisation,
- diminution du nombre de paramètres (ex : LASSO).

Exemple : estimation de densité avec histogramme.

Sélection de modèle

Ainsi, la sélection de modèle ne peut pas se faire uniquement en prenant en compte un écart de vraisemblance. On utilise donc l'un des critères suivants, qu'il faut minimiser :

- AIC (Akaike information criterion) :

$$AIC = 2k - 2 \ln(\hat{L}),$$

où k est le nombre de paramètres du modèle, \hat{L} la vraisemblance la plus forte pour le modèle estimé.

- BIC (Bayesian information criterion) :

$$BIC = \ln(n)k - 2 \ln(\hat{L}),$$

où n est le nombre d'observations.

Ces deux critères ne permettent pas de dire qu'un modèle est bon mais simplement de faire le tri entre plusieurs modèles. Critère à utiliser alors pour quantifier cet écart : **vraisemblance relative**, définie par

$$\rho = \exp \left(\frac{AIC_1 - AIC_2}{2} \right).$$

Le modèle 2 est ρ fois autant vraisemblable que le modèle 1 de minimiser la perte d'information induite par le recours à un modèle.

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 **L'approche statistique**
 - **Régularisation et statistiques non-paramétriques**
 - Surapprentissage et sélection de modèle
 - Méthodes de lissage
 - LASSO
- 5 Exemples de calibrations

Smoothing nonparametric statistics

Smoothing time-dependent (or depending on another parameter, such as a strike, a residual maturity in the framework of model calibration) raw statistics :

- (exponentially weighted) moving average : rule of thumb, thus inaccurate description, lagged predictions, **but** easy to compute.
- decomposition in a series of orthogonal functions (Fourier, wavelets) : many methodological choices (truncation, linear or non-linear filtering of coefficients, mother wavelet), high computational cost (nested integrals), tricky predictions (border effects), **but** accurate description.

The problem, when smoothing the raw statistic \hat{H} :

$$\mathcal{H} = \operatorname{argmin}_{h \in C^2([0,1], \mathbb{R})} \int_0^1 \left[(h(t) - \hat{H}(t))^2 + \lambda h'(t)^2 \right] dt.$$

Fidelity term and smoothness penalty : not that new... e.g. Whittaker-Henderson problem in actuarial literature, or in variational image processing. But always solved (sometimes for more general frameworks) by numerical schemes.

Variational problem

The problem, when smoothing the raw statistic \hat{H} :

$$\mathcal{H} = \operatorname{argmin}_{h \in \mathcal{C}^2([0,1], \mathbb{R})} \int_0^1 \left[(h(t) - \hat{H}(t))^2 + \lambda h'(t)^2 \right] dt. \quad (1)$$

It can be written

$$\mathcal{H} = \operatorname{argmin}_{h \in \mathcal{C}^2([0,1], \mathbb{R})} \int_0^1 \mathcal{L}(t, h(t), h'(t)) dt,$$

so that we can use the Euler-Lagrange equation (explained therefater) :

$$\frac{\partial}{\partial h} \mathcal{L}(t, \mathcal{H}(t), \mathcal{H}'(t)) = \frac{d}{dt} \frac{\partial}{\partial h'} \mathcal{L}(t, \mathcal{H}(t), \mathcal{H}'(t)). \quad (2)$$

In our framework, we have a simple non-homogeneous linear differential equation :

$$\mathcal{H}''(t) - \frac{\mathcal{H}(t)}{\lambda} = -\frac{\hat{H}(t)}{\lambda}. \quad (3)$$

Euler-Lagrange equation

We are looking for h minimizing

$$J(h) = \int_0^1 \mathcal{L}(t, h(t), \dot{h}(t)) dt.$$

Differential calculus provides us with :

$$J(h + \varepsilon) = J(h) + DJ(h, \varepsilon) + o(\varepsilon),$$

for a given function ε such that $\varepsilon(0) = \varepsilon(1) = 0$, and where $DJ(h, \varepsilon)$ is linear in ε . More exactly, $DJ(h, \varepsilon)$ is equal to :

$$DJ(h, \varepsilon) = \int_0^1 \left[\frac{\partial \mathcal{L}}{\partial h}(t, h(t), \dot{h}(t)) \varepsilon + \frac{\partial \mathcal{L}}{\partial \dot{h}}(t, h(t), \dot{h}(t)) \dot{\varepsilon} \right] dt.$$

At the optimum, $DJ(h, \varepsilon) = 0$ and we can integrate by parts the second part of the integral :

$$\int_0^1 \frac{\partial \mathcal{L}}{\partial \dot{h}} \times \frac{d\varepsilon}{dt} dt = \left[\frac{\partial \mathcal{L}}{\partial \dot{h}} \varepsilon \right]_{t=0}^1 - \int_0^1 \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{h}} \times \varepsilon dt,$$

in which the $[\cdot]_{t=0}^1$ is equal to zero because $\varepsilon(0) = \varepsilon(1) = 0$.

Finally :

$$\frac{\partial}{\partial h} \mathcal{L}(t, h(t), \dot{h}(t)) = \frac{d}{dt} \frac{\partial}{\partial \dot{h}} \mathcal{L}(t, h(t), \dot{h}(t)).$$

The solution with two parameters

Théorème

Let $\hat{H} : [0, 1] \rightarrow \mathbb{R}$ be two times differentiable and $\lambda > 0$. Then, a solution $\mathcal{H} : [0, 1] \rightarrow \mathbb{R}$ of equation (1) is necessarily of the form :

$$\mathcal{H} : t \mapsto e^{t/\sqrt{\lambda}} \left(A - \frac{1}{2\sqrt{\lambda}} \int_0^t e^{-s/\sqrt{\lambda}} \hat{H}(s) ds \right) + e^{-t/\sqrt{\lambda}} \left(B + \frac{1}{2\sqrt{\lambda}} \int_0^t e^{s/\sqrt{\lambda}} \hat{H}(s) ds \right),$$

where $A, B \in \mathbb{R}$.

- 1 First (general solution), compute $\Phi \hat{H}$, where Φ is the operator such that, for $h \in \mathcal{C}^2([0, 1], \mathbb{R})$ and $t \in [0, 1]$:

$$\Phi h(t) = -\frac{e^{t/\sqrt{\lambda}}}{2\sqrt{\lambda}} \int_0^t e^{-s/\sqrt{\lambda}} h(s) ds + \frac{e^{-t/\sqrt{\lambda}}}{2\sqrt{\lambda}} \int_0^t e^{s/\sqrt{\lambda}} h(s) ds. \quad (4)$$

- 2 Then (fit to *initial conditions*), project the residual in the functional space generated by $t \mapsto e^{t/\sqrt{\lambda}}$ and $t \mapsto e^{-t/\sqrt{\lambda}}$.

A closed formula in the time series approach

Théorème

Let $\hat{H}(t_1), \dots, \hat{H}(t_n)$ be n statistics associated to times $t_1 = 0 < t_2 < \dots < t_n = 1$, with $n > 1$. Then, the unique function $\mathcal{H} : \{t_0, \dots, t_n\} \rightarrow \mathbb{R}$ such that $\sum_{i=1}^n (\mathcal{H}(t_i) - \hat{H}(t_i))^2 dt$ is minimal and of the form

$$\mathcal{H} : t \mapsto \Phi^d \hat{H}(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}},$$

where Φ^d is the discretized version of the operator introduced in equation (4) and $\lambda > 0$, is characterized by :

$$\begin{pmatrix} A \\ B \end{pmatrix} = \frac{1}{\alpha} \begin{pmatrix} \left(\sum_{i=1}^n e^{-2t_i/\sqrt{\lambda}} \right) \left(\sum_{i=1}^n e^{t_i/\sqrt{\lambda}} [1 - \Phi^d] \hat{H}(t_i) \right) - n \left(\sum_{i=1}^n e^{-t_i/\sqrt{\lambda}} [1 - \Phi^d] \hat{H}(t_i) \right) \\ \left(\sum_{i=1}^n e^{2t_i/\sqrt{\lambda}} \right) \left(\sum_{i=1}^n e^{-t_i/\sqrt{\lambda}} [1 - \Phi^d] \hat{H}(t_i) \right) - n \left(\sum_{i=1}^n e^{t_i/\sqrt{\lambda}} [1 - \Phi^d] \hat{H}(t_i) \right) \end{pmatrix},$$

with $\alpha = \left(\sum_{i=1}^n e^{2t_i/\sqrt{\lambda}} \right) \left(\sum_{i=1}^n e^{-2t_i/\sqrt{\lambda}} \right) - n^2$.

Accuracy of the variational smoother

Théorème

Let Φ be the operator introduced in equation (4) and H an analytic function $[0, 1] \rightarrow \mathbb{R}$. Then, for $\lambda \rightarrow 0$, there exists $\xi \in [0, 1]$ such that :

$$\min_{(A,B) \in \mathbb{R}^2} \int_0^1 \left(\Phi H(t) + Ae^{t/\sqrt{\lambda}} + Be^{-t/\sqrt{\lambda}} - H(t) \right)^2 dt \leq 4\lambda^2 H''(\xi)^2 + o(\lambda^2).$$

In particular, if H is an affine function, the variational smoothing reproduces it without any error. Indeed, as proved by integration by parts in the computation of Φ , if H is an affine function such that $H(t) = a_0 + a_1 t$, then $\Phi H(t) = H(t) - Ae^{t/\sqrt{\lambda}} - Be^{-t/\sqrt{\lambda}}$, with $A = (a_0 + a_1 \sqrt{\lambda})/2$ and $B = (a_0 - a_1 \sqrt{\lambda})/2$.

A simulation study - Metrics for a comparison

- The mean integrated squared error (MISE), average accuracy of the method :

$$R_0 = \frac{1}{N} \sum_{i=1}^N \left[\mathcal{H} \left(\frac{i}{N} \right) - H \left(\frac{i}{N} \right) \right]^2,$$

for N observations in $[0, 1]$, \mathcal{H} being the smooth estimate of H .

- The average bias in $t = 1$, that is on the right of the sample :

$$R_1 = \mathcal{H}(1) - H(1).$$

Ability of the technique not to create artificial gaps at the border of the sample.
Important feature in a predictive framework.

- The average quadratic slope, which determines to which extent the erratic estimate has been smoothed :

$$R_2 = \frac{1}{N-1} \sum_{i=2}^N \left[\frac{\mathcal{H} \left(\frac{i}{N} \right) - \mathcal{H} \left(\frac{i-1}{N} \right)}{1/N} \right]^2.$$

An example : exponential function

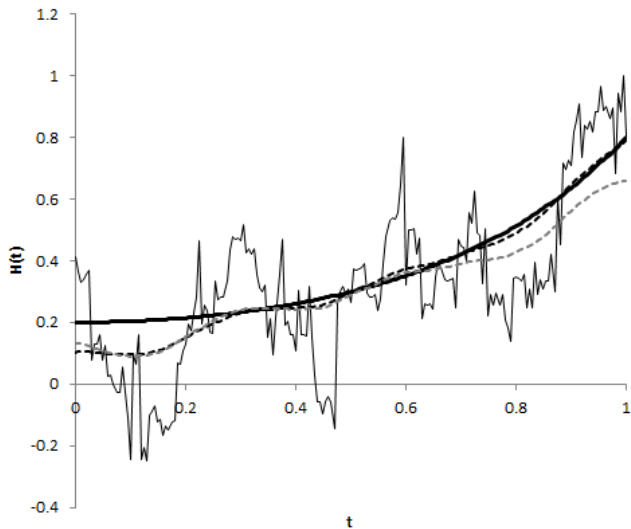
We first simulate a mBm on the time interval $[0, 1]$ with an exponential function in $[0, 1]$ for the Hurst exponent ranging from 0.2 to 0.8 :

$$t \mapsto 0.2 + 0.6 \frac{\exp(t^2) - 1}{\exp(1) - 1}.$$

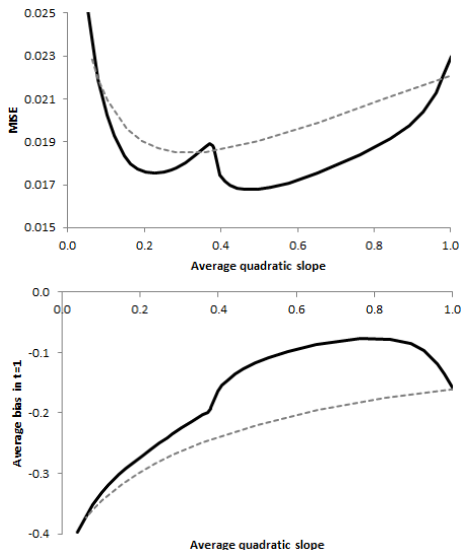
- The time interval is discretized in 200 points.
- Each raw estimate is computed on a window of 25 observations.
- We then obtain the smoothed Hurst exponents by two methods :
 - 1 a variational smoothing of parameter 0.02
 - 2 an exponentially weighted moving average of parameter 0.04.

Next figures : a trajectory, the average metrics for 100 trajectories (black dots : variational smoothing ; grey dots : moving average).

Example : a simulation



Example : average accuracy



- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 **L'approche statistique**
 - **Régularisation et statistiques non-paramétriques**
 - Surapprentissage et sélection de modèle
 - Méthodes de lissage
 - **LASSO**
- 5 Exemples de calibrations

Lasso

LASSO : *Least Absolute Shrinkage and Selection Operator*, introduit par Robert Tibshirani en 1996.

- Alternative à la régression *ridge*, ie régularisation de Tikhonov
 $\operatorname{arginf}_x \|Tx - y\|_2^2 + \lambda \|x\|_2^2$, comme solution du problème inverse mal posé $Tx = y$.
- *Ridge* ne réduit pas le nombre de paramètres et n'améliore donc par BIC ou AIC (le détériore même).

Recall our example: $n = 50$, $p = 30$; true coefficients: 10 are nonzero and pretty big, 20 are zero

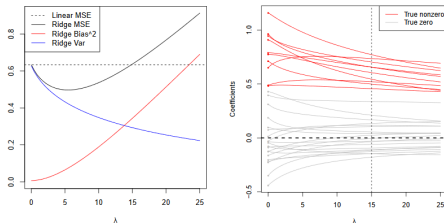


FIGURE – Ridge regression. Source : Tibshirani.

Lasso

- Dans le lasso, la pénalité est une norme 1 et non 2 :

$$\hat{x} = \underset{x \in \mathbb{R}^p}{\operatorname{arginf}} \|Tx - y\|_2^2 + \lambda \|x\|_1,$$

où $\|x\|_1 = \sum_{i=1}^p |x_i|$.

- Simplement grâce à la norme 1, quand λ augmente, de plus en plus de paramètres sont réduits exactement à la valeur 0 : le lasso opère automatiquement une sélection de variables.
- On peut ne pas inclure certains paramètres dans la pénalité si on veut les garder.
- Tous les paramètres ne sont pas forcément à la même échelle (par exemple des prix et des dates). Il faut donc d'abord les convertir pour qu'ils soient comparables.
- Cette sélection des paramètres pertinents permet l'**interprétation** d'un modèle compliqué.

Lasso

Our running example from last time: $n = 50$, $p = 30$, $\sigma^2 = 1$, 10 large true coefficients, 20 small. Here is a visual representation of lasso vs. ridge coefficients (with the same degrees of freedom):

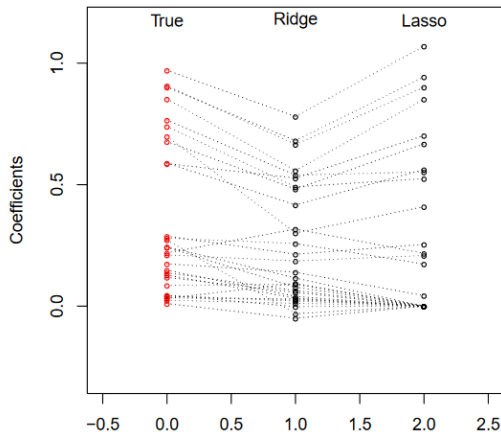


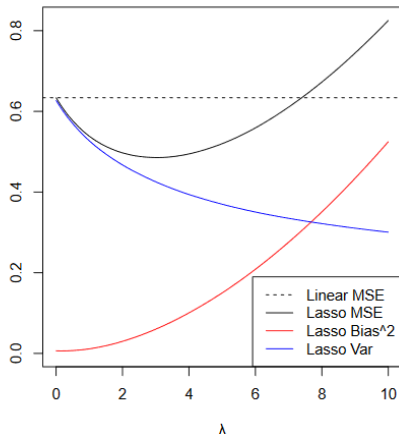
FIGURE — Source : Tibshirani

Calibration

Lasso

Interprétation en termes de lien entre λ et biais/variance similaire à Tikhonov : quand λ augmente, le biais augmente et la variance diminue.

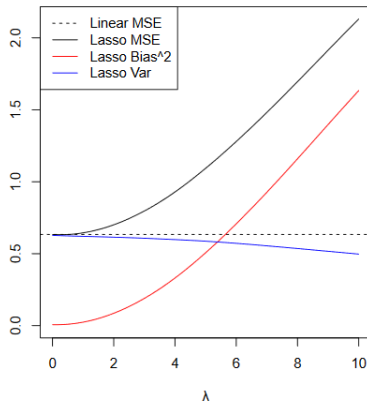
Example: $n = 50$, $p = 30$; true coefficients: 10 large, 20 small



Lasso

Ne fonctionne pas dans tous les cas, notamment dans ceux où parcimonie pas de mise.

Example: $n = 50$, $p = 30$; true coefficients: 30 moderately large



Note that here, as opposed to ridge regression the variance doesn't decrease fast enough to make the lasso favorable for small λ

Lasso

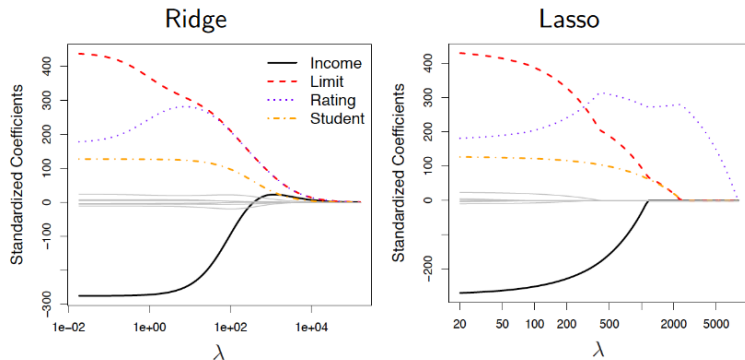


FIGURE – Variation des paramètres standardisés dans cas ridge/lasso. Source : Tibshirani.

Lasso

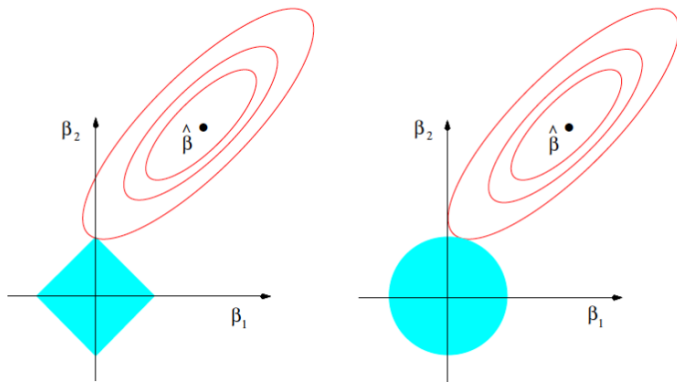


FIGURE – Pourquoi le lasso réduit-il des paramètres à zéro ? C'est une question de normes (1 à gauche, 2 à droite). Le problème peut être vu comme optimisation sous contrainte : $\hat{x} = \underset{x \in \mathbb{R}^p}{\operatorname{arginf}} \|Tx - y\|_2^2$ sous contrainte que $\|x\|_1 \leq \theta$ (ou $\|x\|_2^2 \leq \theta$ pour ridge). Quand la contrainte est saturée, à cause du pic sur la norme, l'un des deux paramètres vaut zéro. Source : Tibshirani.

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 L'approche statistique
- 5 Exemples de calibrations
 - Courbe de taux
 - Calibration
 - Régularisations paramétriques

Taux zéro coupon et taux actuariel

Définition

Obligation a flux C_i aux instants t_i . Le prix observé est P_t . Le **taux actuariel**, R , est tel que :

$$P_t = \sum_{i=1}^N \frac{C_i}{(1 + R)^{t_i - t}}.$$

Définition

Si l'obligation a un seul flux, à l'instant T , son prix est : $P_t = C / (1 + R^{ZC})^{T - t}$. Le taux R^{ZC} est un **taux zéro-coupon** (c'est le **taux actuariel** de ce titre).

Calibration d'une courbe de taux, deux méthodes possibles :

- On peut donc actualiser chaque flux d'un titre au taux zéro-coupon de la maturité correspondante (courbe calibrée sur un univers d'obligations) : seul le dernier flux sera actualisé à un taux spécifique.
- On peut actualiser tous les flux au même taux actuariel : ce taux dépend entièrement du titre.

Ainsi, seule l'actualisation au taux ZC est cohérente entre les différents titres.

Taux zéro coupon et taux actuariel

Valorisation

Une fois que la courbe ZC (pas la courbe de taux actuariel) est calibrée, on peut valoriser n'importe quel titre (ayant servi ou non à calibrer la courbe) :

$$P_t = \sum_{i=1}^N \frac{C_i}{(1 + R_{t_i}^{ZC})^{t_i - t}}.$$

- Si la courbe de taux est croissante, le taux actuariel (moyenne de ZC de maturités inférieure à maturité du titre) sera inférieur au taux ZC.
- On peut vouloir avoir une telle courbe dans différents cadres de valorisation :
 - calibrée sur un univers liquide d'obligations pour valoriser des titres illiquides,
 - calibrée sur un univers représentant un monde *sans risque* (taux gouvernemental, de certains pays) pour valoriser des obligations d'entreprises (en ajoutant un spread au dénominateur ou une probabilité de défaut au numérateur).

Calibration de courbe zéro coupon : *bootstrap* 1/2

La courbe est construite à partir d'un univers d'obligations (ayant ou non des coupons non-nuls).

1 On trie les titres par maturité croissante.

2 Pour le titre de plus faible maturité :

- s'il ne paye qu'un flux C , en date T , le prix du jour est $P_0 = C / (1 + R_{0,T})^T$, ce qui nous permet d'écrire :

$$R_{0,T} = \left[\frac{C}{P_0} \right]^{1/T} - 1.$$

- s'il verse un coupon intermédiaire, en l'absence de données sur le début de la courbe, on considère que le taux est constant avant la date de maturité, égal à $R_{0,T}$. Ainsi, on a :

$$P_0 = \sum_{i=1}^N \frac{C_i}{(1 + R_{0,T})^{t_i}}.$$

Résolution par optimisation numérique. Initialisation obtenue avec le développement limité à l'ordre 1, $P_0 = \sum_{i=1}^N C_i \times (1 - t_i R_{0,T}) + o(R_{0,T})$, ce qui donne l'approximation suivante :

$$R_{0,T} \approx \frac{\left[\sum_{i=1}^N C_i \right] - P_0}{\sum_{i=1}^N C_i t_i}.$$

Calibration de courbe zéro coupon : *bootstrap* 2/2

...

- 3 Après les j premiers titres, on déduit le taux en la date de maturité du $(j + 1)$ -ème titre à partir des résultats déjà obtenus : on interpole la courbe de façon linéaire entre toutes les dates de maturité connues et antérieures à T_j , et on suppose que le taux est constant et égal à R_{0,T_1} avant la maturité T_1 . Le prix du titre considéré est alors

$$P_0^{j+1} = \sum_{i=1}^{N_{j+1}} \frac{C_i^{j+1}}{(1 + R_{0,t_i})^{t_i}}$$

et on complète la courbe avec :

$$R_{0,T_{j+1}} = \left[\frac{C_{N_{j+1}}}{P_0 - \sum_{i < N_{j+1}} \frac{C_i}{(1 + R_{0,t_i})^{t_i}}} \right]^{1/T_{j+1}} - 1.$$

- 1 Algorithmes d'optimisation
- 2 Calibration de volatilité implicite
- 3 Théorie de la calibration
- 4 L'approche statistique
- 5 Exemples de calibrations**
 - Courbe de taux
 - Calibration
 - Régularisations paramétriques

Splines cubiques (1/3)

Courbe empirique inexploitable si présence d'*outliers*.

- Méthode d'interpolation utilisant des polynômes d'interpolation de degré 3, raccordés en des points nommés **nœuds**.
- K nœuds : $x_0 < \dots < x_{K-1}$.
- Splines s'écrivent donc

$$P(x) = \sum_{k=1}^{K-1} P_k(x) \mathbb{1}_{x \in [x_{k-1}, x_k]},$$

où $P_k(x) = a_k(x - x_{k-1})^3 + b_k(x - x_{k-1})^2 + c_k(x - x_{k-1}) + d_k$.

- Nombre de paramètres : $4(K - 2)$. Les trois quarts sont imposés par le modèle.
- Conditions imposées :
 - continuité ($K - 2$ paramètres fixés) : $\forall k > 1$, on a $P_k(x_{k-1}) = P_{k-1}(x_{k-1})$,
 - continuité de la dérivée ($K - 2$ paramètres fixés) : $\forall k > 1$, on a $P'_k(x_{k-1}) = P'_{k-1}(x_{k-1})$,
 - continuité de la dérivée seconde ($K - 2$ paramètres fixés) : $\forall k > 1$, on a $P''_k(x_{k-1}) = P''_{k-1}(x_{k-1})$.

Splines cubiques (2/3)

- Concrètement :
 - entre les deux premiers nœuds, estimation d'un polynôme de degré 3 (ou 2 si on suppose le taux court connu, que l'on met dans le paramètre d_1), ce qui est un problème linéaire de dimension 3.
 - deuxième intervalle (et identique ensuite) : on fixe d_2 , c_2 et b_2 par les trois conditions de continuité, et on est donc réduit à un problème linéaire de dimension 1 pour trouver a_2 (taux empiriques : $R(t_{1_k}), \dots, R(t_{n_k})$ pour $t_{1_k}, \dots, t_{n_k} \in [x_{k-1}, x_k]$) :

$$\operatorname{argmin}_{a_2} \frac{1}{n_k} \sum_{i=1_k}^{n_k} (R(t_i) - P_k(t_i))^2.$$

- Nombre de nœuds : recommandé moins de 8, ou quantité comparable à Svensson (6).
- Choix des nœuds : même nombre de points sur la courbe empirique entre deux nœuds.
- Application de ce modèle à la courbe *forward* ou *spot* ; le faire sur le *forward* permet d'avoir une courbe *spot* plus régulière, par intégration.

Splines cubiques (3/3)

Résultat en général oscillant, donc ajout de pénalisation :

$$\operatorname{argmin}_{a_2} \frac{1}{n_k} \sum_{i=1_k}^{n_k} (R(t_i) - P_k(t_i))^2 + \lambda \int_{x_{k-1}}^{x_k} P_k''(x)^2 dx.$$

Mais le lissage devient alors trop efficace à faible maturité et efface des bosses, donc il faut utiliser un λ adaptatif, faible à courte maturité et grand à plus grande maturité :

$$\operatorname{argmin}_{a_2} \frac{1}{n_k} \sum_{i=1_k}^{n_k} (R(t_i) - P_k(t_i))^2 + \int_{x_{k-1}}^{x_k} \lambda(x) P_k''(x)^2 dx.$$

Modèle de Nelson-Siegel

Nelson et Siegel observent que les taux *forward* semblent dériver de fonctions exponentielles :

$$f(0, t) = \beta_0 + \beta_1 \exp\left(\frac{-t}{\tau}\right) + \beta_2 \frac{t}{\tau} \exp\left(\frac{-t}{\tau}\right).$$

Le taux *spot* s'écrit alors :

$$\begin{aligned} R_{0,\theta} &= \beta_0 \\ &+ \beta_1 \left[\frac{1 - \exp\left(-\frac{\theta}{\tau_{SM}}\right)}{\frac{\theta}{\tau_{SM}}} \right] \\ &+ \beta_2 \left[\frac{1 - \exp\left(-\frac{\theta}{\tau_{SM}}\right)}{\frac{\theta}{\tau_{SM}}} - \exp\left(-\frac{\theta}{\tau_{SM}}\right) \right] \end{aligned}$$

avec :

- $\beta_0 = \lim_{\theta \rightarrow \infty} R_{0,\theta}$ et s'interprète donc comme le taux de long terme ;
- $\beta_1 = \lim_{\theta \rightarrow 0} R_{0,\theta} - \beta_0$ et s'interprète donc comme le *spread long/short* ; par conséquent, $\beta_0 + \beta_1$ correspond au taux *overnight* ;
- β_2 s'interprète comme un paramètre de courbure ;
- $\tau_{SM} > 0$ paramètre d'échelle qui mesure le taux de *decay* du court et moyen terme vers 0.

Modèle de Svensson

Le modèle de Svensson, amélioration du modèle de Nelson et Siegel pour prendre en compte une éventuelle bosse suivie d'une creux sur la courbe, dont le taux zéro-coupon de composition continue en 0 de maturité θ est :

$$\begin{aligned}
 R_{0,\theta} = & \beta_0 \\
 & + \beta_1 \left[\frac{1 - \exp(-\frac{\theta}{\tau_{SM}})}{\frac{\theta}{\tau_{SM}}} \right] \\
 & + \beta_2 \left[\frac{1 - \exp(-\frac{\theta}{\tau_{SM}})}{\frac{\theta}{\tau_{SM}}} - \exp(-\frac{\theta}{\tau_{SM}}) \right] \\
 & + \beta_3 \left[\frac{1 - \exp(-\frac{\theta}{\tau_S})}{\frac{\theta}{\tau_S}} - \exp(-\frac{\theta}{\tau_S}) \right]
 \end{aligned}$$

avec, β_3 correspondant à un paramètre de courbure de la partie courte de la courbe et $\tau_S > 0$ le paramètre d'échelle correspondant.