

A1.1  
6  
796

Estimation non paramétrique d'une fonction

d'évaluation d'options avec contraintes de

monotonie et de convexité

\*\*\*\*\*

Fatima-Zahra NQI

\*\*\*\*\*

Directeur du projet: René Garcia

\*\*\*\*\*

Université de Montréal et CIRANO

\*\*\*\*\*

Rapport de Maîtrise en Sciences Économiques

\*\*\*\*\*

24 Mars 1999

\*\*\*\*\*

CENTRE DE DOCUMENTATION  
à 5 111 111  
SCIENCES ÉCONOMIQUES U DE M

## TABLE DE MATIÈRES

1. Introduction
2. Modélisation et estimation du prix d'option
  - 2.1. Approche paramétrique: Black-Scholes
  - 2.2. Approche non paramétrique
  - 2.3. Estimation par les réseaux de neurones
  - 2.4. Estimation par les noyaux
3. Application de l'isotonisation aux estimateurs
  - 3.1. Définitions
  - 3.2. Estimation par réseaux de neurones isotonisés
  - 3.3. Estimation par les noyaux isotonisés
4. Application numérique
  - 4.1. Données
  - 4.2. Simulations et comparaisons
  - 4.3. Algorithme pour l'isotonisation
5. Conclusion

## ANNEXE MATHÉMATIQUE

1. Introduction
2. Énoncé du problème
3. Rappels de quelques propriétés d'isotonisation
4. Estimation par réseaux de neurones et isotonisation
  - 4.1. Estimation par réseaux de neurones ordinaires
  - 4.2. L'isotonisation de l'estimateur  $m_R$
  - 4.3. L'isotonisation de la dérivée de  $m_{RI}$
  - 4.4. Définition de l'estimateur  $\hat{m}_R$
  - 4.5. Propriétés de l'estimateur  $\hat{m}_R$
5. Estimation par les noyaux et isotonisation
  - 5.1. Estimation par les noyaux ordinaires
  - 5.2. L'isotonisation de  $m_S$
  - 5.3. L'isotonisation de la dérivée de  $m_{SI}$
  - 5.4. Définition de l'estimateur  $\hat{m}_S$
  - 5.5. Propriétés de l'estimateur  $\hat{m}_S$

## Résumé

Dans ce travail, on veut estimer d'une façon non paramétrique la fonction d'évaluation d'options lorsqu'on lui impose des contraintes de monotonie et de convexité par la méthode d'isotonisation. Pour satisfaire la contrainte de monotonie, la régression isotonique consiste à projeter la fonction non paramétrique dans l'ensemble des fonctions monotones croissantes. Pour satisfaire la contrainte de convexité, la même opération est appliquée à la dérivée première de cette fonction. Dans l'estimation non paramétrique de la dérivée de la fonction d'évaluation, il est souvent supposé que le prix d'option est une fonction homogène de degré un par rapport au prix  $S_t$  de l'actif sous-jacent et au prix d'exercice  $K$ . Cette homogénéité implique la convexité de la fonction. Bien que des résultats théoriques établissent que les estimateurs non paramétriques approximent bien la fonction et ses dérivées, il est souvent difficile d'y parvenir en pratique où les conditions asymptotiques ne sont pas réalisables pour un échantillon fini. Avec des données simulées et des données réelles d'options sur l'indice S&P 500, on montrera que les contraintes de monotonie et de convexité nous permettent d'estimer avec plus de précision la fonction de prix d'option et le ratio de couverture.

**Mots clés:** Prix d'option, estimation non paramétrique, réseaux de neurones, isotonisation.

## 1 Introduction.

Une série d'articles (Huchinson et al. (1994), Gouriéroux, Monfort et Tenreiro (1995), Ait-Sahalia et Lo (1997), Garcia et Gençay (1998), Broadie et al. (1998)) ont utilisé les méthodes non paramétriques pour estimer la formule d'évaluation d'options. Puisque les méthodes d'évaluation non paramétriques n'imposent aucune dynamique sur le prix de l'actif sous-jacent, elles s'avèrent plus robustes que les méthodes paramétriques. La plupart du temps, on suppose que la fonction d'évaluation d'option est homogène de degré un<sup>1</sup> par rapport au prix de l'actif sous-jacent et au prix d'exercice. En divisant tous les termes par le prix d'exercice, on réduira le nombre de variables dans l'apprentissage de la fonction d'évaluation non paramétrique. Ceci représente un avantage puisque le taux de convergence des estimateurs non paramétriques ralentit considérablement lorsque le nombre de variables augmente. Broadie et al. (1996a), qui a aussi utilisé les méthodes non paramétriques pour évaluer les options, montre qu'une autre raison pour justifier l'homogénéité est la non stationnarité des prix d'options et des prix d'actions. L'hypothèse d'homogénéité n'est pas consistante avec la dynamique de l'ensemble des prix. Merton (1973) montre que l'indépendance sérielle de l'ensemble des rendements pour des données générées est une condition suffisante pour l'homogénéité. Dans un contexte de non arbitrage, Garcia et Renault (1995) établissent que l'indépendance conditionnelle sous la mesure de probabilité des prix entre les rendements futurs et le prix courant est une condition nécessaire et suffisante pour l'homogénéité de la fonction d'évaluation des prix. L'homogénéité de la fonction d'évaluation des options implique la convexité. Il paraît naturel de supposer que le prix d'option à la date  $t$  possède la propriété de convexité par rapport au prix terminal de l'actif sous-jacent.

Les contraintes de monotonie croissante et de convexité sont satisfaites dans le modèle classique de Black-Scholes. Mais comme ce modèle impose une dynamique bien précise sur le prix de l'actif sous-jacent, on s'intéressera plutôt à des méthodes non paramétriques qui n'imposent aucune structure sur ces prix. Parmi ces méthodes les plus utilisées, il y a la méthode d'apprentissage de la fonction d'évaluation en utilisant les réseaux de neurones et la méthode d'estimation par les noyaux.

Dans le cas des réseaux de neurones, la propriété de monotonie croissante est préservée, par contre, ce n'est pas le cas numériquement pour la propriété

---

<sup>1</sup>Soit  $f$  une fonction définie de  $\mathbf{R}^n$  dans  $\mathbf{R}^p$ . On dit que  $f$  est homogène de degré un par rapport à la  $j$ ème variable si et seulement si  $\forall \alpha > 0$ , on a  $f(x_1, \dots, \alpha x_j, \dots, x_n) = \alpha f(x_1, \dots, x_j, \dots, x_n)$ .

de convexité. Il existe des résultats théoriques (voir Gallant et White (1992); Hornik, Stinchcombe et White (1989) et (1990); Hornik (1991)) qui montrent que les dérivées des estimateurs par les réseaux de neurones approchent bien les dérivées de la fonction d'évaluation lorsque le nombre de neurones augmente asymptotiquement avec le nombre de données. Malheureusement, ceci est difficilement réalisable numériquement en échantillon fini. On peut déduire que numériquement, la propriété de convexité n'est pas garantie par ce type d'estimateurs. Dans ce cas, on doit introduire une méthode comme l'isotonisation, qui par ses propriétés, préserve théoriquement et numériquement, sans condition asymptotique, les propriétés de monotonie et de convexité.

La procédure d'isotonisation d'une fonction consiste à la projeter sur l'ensemble des fonctions monotones croissantes. Par ce moyen, la propriété de monotonie est satisfaite. Pour satisfaire la propriété de convexité, il faut appliquer la même opération à la dérivée première de la fonction.

En appliquant doublement l'opération d'isotonisation aux réseaux de neurones, on obtient un nouvel estimateur qui est monotone croissante et convexe.

Par ailleurs, il existe dans la littérature d'autres techniques qui permettent de préserver la monotonie et la convexité dans l'estimation. La plus utilisée de ces techniques est la technique des splines cubiques appliquées aux estimateurs à noyaux, pour cela, il faut voir les travaux d'Utreras (1985), Ramsey (1988), Schmidt et Scholz (1990), Gaylord et Ramirez (1991), Matzkin (1991), Schwetlick et Kunert (1993), Tantiyaswasdikul et Woodroffe (1994), Matzkin (1994), Mammen et Thomas-Agnan (1996).

L'un des avantages dans l'utilisation de la méthode de l'isotonisation est qu'elle applicable à n'importe quelle estimateur et pas uniquement à l'estimateur à noyaux comme le cas de la méthode précédente.

Dans la première section, on expliquera les étapes suivies pour modéliser le prix d'option en fonction du prix de l'actif sous-jacent, du prix d'exercice et de la maturité. On expliquera les raisons pour lesquelles on choisit de travailler avec des modèles non paramétriques plutôt que d'utiliser les modèles paramétriques. On donnera aussi une description de quelques méthodes d'estimation non paramétriques. On y donnera aussi les arguments nécessaires pour réduire le nombre de variables.

Dans la deuxième section, on donnera la définition de la procédure d'isotonisation et la manière avec laquelle elle a été appliquée aux estimateurs par réseaux de neurones et à noyaux. On expliquera ses avantages concernant la satisfaction des contraintes de monotonie et de convexité dans l'estimation.

Dans la troisième section, on appliquera numériquement cette opération d'isotonisation mais uniquement dans le cas d'estimation par les réseaux de neurones. On expliquera les données utilisées et on fera des comparaisons entre l'estimation avec et sans isotonisation pour mettre en évidence les gains qu'on obtient surtout au niveau du ratio de couverture à la suite de l'application de cette opération.

À la fin, une annexe mathématique explique en détail l'opération d'isotonisation et son application aux estimateurs non paramétriques cités auparavant.

## 2 Modélisation et estimation du prix d'option et du ratio de couverture.

Il existe deux approches pour évaluer le prix d'option et le ratio de couverture en fonction du prix de l'actif sous-jacent:

- l'approche traditionnelle dite paramétrique qui impose une certaine structure sur le prix  $S_t$  de l'actif sous-jacent et suppose qu'il doit être complètement déterminé grâce à un nombre fini de paramètres;
- l'approche moderne dite non paramétrique ne précise pas explicitement le processus suivi par le prix de l'actif, mais permet de le déduire à partir des données soumises à certaines conditions de régularité.

### 2.1 Approche paramétrique: Black-Scholes.

Le modèle le plus utilisé pour exprimer le prix  $C_t$  d'une option européenne d'achat en fonction du prix  $S_t$  de l'actif sous-jacent, du prix d'exercice  $K$  et de la maturité  $\tau = T - t$  est le modèle de Black-Scholes. Ce modèle suppose que la série  $\{S_t, t = 1, \dots, T\}$  suit un mouvement brownien géométrique avec une moyenne  $\mu$  et une volatilité  $\sigma$ , plus précisément:

$$(2.1) \quad \forall t = 1, \dots, T \quad dS_t = \mu S_t dt + \sigma S_t dW_t,$$

où  $W_t$  est la variable aléatoire du processus brownien. En pratique, on peut générer cette série à l'aide d'une variable aléatoire  $Z_i$  distribuée normalement avec une moyenne égale à  $\mu/T$  et une variance égale à  $\sigma^2/T$ , soit

$$(2.2) \quad \forall t, \quad S_t = S_0 \exp \left( \sum_{i=1}^t Z_i \right),$$

où  $S_0$  est le prix initial de l'actif sous-jacent. Lorsque la moyenne  $\mu$  est égale au taux sans risque  $r$ , la forme la plus simple du modèle de Black-Scholes s'écrit de la façon suivante:

$$(2.3) \quad C_t = S_t \Phi(d_1(S_t)) - K e^{-r(T-t)} \Phi(d_2(S_t)),$$

où  $\Phi$  désigne la fonction de distribution normale cumulative, et où  $d_1$  et  $d_2$  sont des fonctions de  $S_t$ , plus précisément:

$$(2.4) \quad d_1(S_t) = \frac{\ln(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}},$$

et

$$(2.5) \quad d_2(S_t) = d_1(S_t) - \sigma\sqrt{T-t}.$$

Le ratio de couverture  $\Delta_t$  est défini comme la dérivée de  $C_t$  par rapport à  $S_t$ , soit

$$(2.6) \quad \Delta_t = \Phi(d_1(S_t)) + S_t \frac{\partial d_1(S_t)}{\partial S_t} \Phi'(d_1(S_t)) - K e^{-r(T-t)} \frac{\partial d_2(S_t)}{\partial S_t} \Phi'(d_2(S_t)),$$

où  $\Phi'$  désigne la dérivée de  $\Phi$ . Sachant que

$$(2.7) \quad \frac{\partial d_1(S_t)}{\partial S_t} = \frac{\partial d_2(S_t)}{\partial S_t} = \frac{1}{S_t \sigma \sqrt{T-t}},$$

on obtient que

$$(2.8) \quad \Delta_t = \Phi(d_1(S_t)) + \frac{S_t \Phi'(d_1(S_t)) - K e^{-r(T-t)} \Phi'(d_2(S_t))}{S_t \sigma \sqrt{T-t}}.$$

Le deuxième terme est nul. En effet, sachant que

$$\forall x, \quad \Phi'(x) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right),$$

et grâce à la relation (2.5), on trouve que

$$S_t \Phi'(d_1(S_t)) - K e^{-r\tau} \Phi'(d_2(S_t)) = \frac{1}{2\pi} \exp\left(-\frac{d_1^2}{2}\right) \left[ S_t - K e^{-r\tau} \exp\left(\frac{-\sigma^2\tau}{2} + d_1\sigma\sqrt{\tau}\right) \right].$$

Sachant que

$$d_1\sigma\sqrt{\tau} = \ln\frac{S_t}{K} + (r + \sigma^2/2)\tau,$$

on trouve que

$$\Delta_t = \Phi(d_1(S_t)).$$

## 2.2 Approche non paramétrique.

Comme tout modèle paramétrique, le modèle de Black-Scholes a des limites puisqu'il impose une structure bien précise sur le prix  $S_t$  et suppose qu'il doit être complètement déterminé grâce à un nombre fini de paramètres. Pour remédier à cela, il existe une autre approche dite non paramétrique et qui ne précise pas explicitement le processus suivi par le prix. Donc une façon naturelle pour évaluer une option européenne d'achat (voir Garcia et Genay (1998)) est d'exprimer son prix  $C_t$  comme une fonction non paramétrique de l'ensemble des variables qui caractérisent cette option, à savoir le prix de l'actif  $S_t$ , le prix d'exercice  $K$  et la maturité  $\tau = T - t$ . On peut écrire alors que

$$(2.9) \quad C_t = f(S_t, K, T - t),$$

où  $f$  est une fonction non-linéaire, suffisamment régulière monotone croissante et convexe par rapport à la première variable. Cette approche a été suivie par Hutchinson, Lo et Poggio (1994). En général, il est très difficile d'estimer la fonction  $f$  qui intervient dans un modèle non paramétrique lorsque le nombre d'inputs est grand. Parmi les méthodes les plus utilisées pour estimer cette fonction, il y a la méthode des réseaux de neurones. Hutchinson et al. (1994) ont supposé que la formule des réseaux pour le prix d'option est homogène de degré un par rapport au prix de l'actif sous-jacent et au prix d'exercice. Cette hypothèse nous permettra de réduire le nombre d'inputs. Cette hypothèse d'homogénéité n'est pas cohérente avec n'importe quelle dynamique du prix de l'actif sous-jacent. Merton (1973) a montré que l'indépendance sérielle des rendements des actifs pour des données qui génèrent le processus est une condition suffisante pour l'homogénéité. Dans un contexte de non-arbitrage, Garcia et Renault (1995) ont établi que l'indépendance conditionnelle sous la mesure de probabilité d'évaluation entre les rendements futurs et le prix courant est une condition nécessaire et suffisante pour l'homogénéité de la fonction d'évaluation d'option. Pour réduire le nombre d'inputs, Hutchinson, Lo et Poggio (1994) divisent le prix de l'option et les arguments de la fonction d'évaluation par  $K$ , ce qui implique à partir du modèle (2.9):

$$(2.10) \quad \frac{C_t}{K} = f\left(\frac{S_t}{K}, 1, \tau\right).$$

Cette forme suppose l'homogénéité de degré un de  $f$  par rapport au prix de l'actif et au prix d'exercice.



Si on se place dans un ensemble d'options européennes d'achat pour lesquelles la maturité  $\tau$  est la même, à partir du modèle (2.10) il existe une fonction  $m$  non-linéaire et suffisamment régulière telle que:

$$(2.11) \quad \frac{C_t}{K} = m\left(\frac{S_t}{K}\right).$$

Dans ce cas le ratio de couverture  $\Delta_t$  qui est défini comme la variation du prix d'option par rapport au prix de l'actif sous-jacent, s'exprime de la façon suivante:

$$(2.12) \quad \Delta_t = \frac{\partial C_t}{\partial S_t} = m'\left(\frac{S_t}{K}\right),$$

où  $m'$  désigne la dérivée première de la fonction  $m$ . Il faut remarquer que l'homogénéité de degré un de la fonction  $f$  par rapport à ses deux premiers arguments implique l'homogénéité de la fonction  $m$ . Sachant que toute fonction homogène de degré un est convexe et que le prix d'option augmente avec le prix de l'actif sous-jacent, on peut déduire que la fonction  $m$  est monotone croissante et convexe.

### 2.3 Estimation par les réseaux de neurones.

Soit  $f(x, \theta)$  une régression, où  $x$  sont des variables et  $\theta$  le vecteur des paramètres. La fonction  $f$  permet de définir le lien entre les variables et les paramètres. L'architecture des réseaux qu'on va utiliser peut être décrite d'une façon très simple: chaque input  $x_i$  envoie un signal à chaque noeud  $j$ , cela implique que chaque noeud  $j$  reçoit une combinaison linéaire des inputs  $x_i$  affectés des poids  $\gamma_{ij}$ , soit la somme:

$$(2.13) \quad \gamma_{j0} + \sum_{i=1}^p \gamma_{ji}x_i,$$

où le terme  $\gamma_{j0}$  représente le biais. L'étape suivante est que le neurone  $j$  envoie un output  $h_j$  tel que:

$$(2.14) \quad h_j = G\left(\gamma_{j0} + \sum_{i=1}^p \gamma_{ji}x_i\right),$$

où  $G$  est la fonction d'activation donnée par

$$(2.15) \quad G(x) = \frac{1}{1 + e^{-\alpha x}} \quad \text{avec} \quad \alpha > 0.$$

$G$  est une fonction logistique qui a les propriétés d'une fonction sigmoïdale, à savoir qu'elle est définie de l'ensemble des réels  $\mathbf{R}$  vers l'intervalle  $[0, 1]$ , qu'elle est monotone croissante, que  $G(x)$  tend vers zéro lorsque  $x$  tend vers  $-\infty$  et que  $G(x)$  tend vers 1 lorsque  $x$  tend vers  $+\infty$ . Les signaux des noeuds  $j = 1, \dots, d$  seront envoyés aux outputs de la même façon que les inputs ont été connectés aux noeuds, c'est à dire une combinaison linéaire des  $h_j$ , donc l'output reçoit  $\beta_0 + \sum_{j=1}^d \beta_j h_j$ . Si on utilise l'expression de  $h_j$  donnée précédemment, on trouve que:

$$(2.16) \quad f(x, \theta) = \beta_0 + \sum_{j=1}^d \beta_j G(\gamma_{j0} + \sum_{i=1}^p \gamma_{ji} x_i).$$

Plusieurs auteurs se sont intéressés aux propriétés de l'approximation universelle pour les réseaux de neurones (Gallant et White (1988,1992); Cybenko (1989); Funahashi (1989); Hornik, Stinchcombe et White (1989,1990)). Certains d'entre eux ont établi que les réseaux de neurones sont capables d'approximer toute fonction mesurable au sens de Borel définie d'un ensemble dans un espace de dimension finie vers un autre avec n'importe quelle degré de précision désiré. On en déduit que les réseaux de neurones représentent une classe d'approximateurs universels. Certains auteurs ont donné des conditions nécessaires et suffisantes sous lesquelles les réseaux de neurones puissent approcher convenablement n'importe quelle fonction et ses dérivées<sup>2</sup>. Ce résultat a été généralisé même pour les fonctions qui ne sont pas différentiables au sens fort<sup>3</sup>, mais qui possèdent des dérivées au sens faible<sup>4</sup>, par exemple les fonctions différentiables par morceaux. Dans notre étude, on utilisera le cas où la fonction d'activation est sigmoïdale. Il faut remarquer qu'il existe aussi des résultats théoriques concernant la fiabilité des réseaux de neurones même lorsque cette fonction n'est pas du type sigmoïdal. On peut citer par exemple H. Hornik (1990) qui a montré que lorsque la fonction d'activation est non constante, bornée et parfois continue, les réseaux de neurones associés sont des estimateurs universels par rapport au critère de la performance  $L^p$ . Gallant et White (1992) ont montré que la classe des réseaux de neurones est dense dans des espaces de Sobolev<sup>5</sup>, ce qui leur associe le caractère de régularité. Si on note  $C_t^R$  et  $\Delta_t^R$  le prix

<sup>2</sup>voir les papiers de Gallant et White (1992); Hornik, Stinchcombe et White (1990).

<sup>3</sup>ie sûrement selon la probabilité de la mesure utilisée. Dans ce cas, il s'agit de la mesure de Borel.

<sup>4</sup>ie presque partout selon la mesure de Borel. On peut dire aussi au sens des distributions.

<sup>5</sup>pour la notion de sobolev, voir papier de Gallant et White (1992).

d'option et le ratio de couverture calculés par la méthode des réseaux de neurones, alors il existe  $\beta_0, \beta_j, \gamma_j$  et  $\theta_j$  pour  $j = 1, \dots, d$  tels que

$$(2.17) \quad \frac{C_t^R}{K} = m_R \left( \frac{S_t}{K} \right) = \beta_0 + \sum_{j=1}^d \beta_j G \left( \gamma_j + \theta_j \frac{S_t}{K} \right),$$

et

$$(2.18) \quad \Delta_t^R = \sum_{j=1}^d \beta_j \theta_j G' \left( \gamma_j + \theta_j \frac{S_t}{K} \right),$$

où  $\beta_0, \beta_j, \gamma_j$  et  $\theta_j$  sont les solutions du problème d'optimisation suivant:

$$(2.19) \quad \min_{\tilde{\beta}_0, \tilde{\beta}_j, \tilde{\gamma}_j, \tilde{\theta}_j} \sum_{i=1}^N \left[ \frac{C_t}{K} - \tilde{\beta}_0 - \sum_{j=1}^d \tilde{\beta}_j G \left( \tilde{\gamma}_j + \tilde{\theta}_j \frac{S_t}{K} \right) \right]^2,$$

où  $(C_t)_t, (S_t)_t$  sont des données.

#### 2.4 Estimation par les noyaux.

Dans cette partie, on va utiliser les estimateurs à noyaux pour approximer les prix d'option et les ratios de couverture correspondants. On choisira le noyau Gaussien  $K_h$  qui est utilisé dans la plupart des cas, il est défini par

$$(2.20) \quad K_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}.$$

On note  $C_t^S$  et  $\Delta_t^S$  le prix d'option et le ratio de couverture calculés par les estimateurs à noyaux (Gaussiens). Ils s'écrivent

$$(2.21) \quad \frac{C_t^S}{K} = m_S \left( \frac{S_t}{K} \right) = \frac{\sum_{i=1}^T K_h \left( \frac{S_t}{K} - \frac{S_i}{K} \right) \frac{C_i}{K}}{\sum_{i=1}^T K_h \left( \frac{S_t}{K} - \frac{S_i}{K} \right)},$$

et

$$(2.22) \quad \Delta_t^S = \frac{\mathcal{A}(S_t) - \mathcal{B}(S_t)}{\left( \sum_{i=1}^T K_h \left( \frac{S_t}{K} - \frac{S_i}{K} \right) \right)^2},$$

avec

$$(2.23) \quad \mathcal{A}(S_t) = \left( \sum_{i \neq t} K_h' \left( \frac{S_t}{K} - \frac{S_i}{K} \right) \frac{C_i}{K^2} \right) \left( \sum_{i=1}^T K_h \left( \frac{S_t}{K} - \frac{S_i}{K} \right) \right),$$

et

$$(2.24) \quad \mathcal{B}(S_t) = \left( \sum_{i=1}^T K_h \left( \frac{S_t - S_i}{K} \right) \frac{C_i}{K} \right) \left( \sum_{i \neq t} K'_h \left( \frac{S_t - S_i}{K} \right) \frac{1}{K} \right).$$

où  $K'_h$  désigne la dérivée du noyau  $K_h$ .

### 3 Application de l'isotonisation aux estimateurs.

#### 3.1 Définitions.

Soit  $\hat{m}$  un estimateur d'une fonction non paramétrique qui est monotone croissante. On note  $\hat{m}_I$  l'estimateur obtenu par l'isotonisation de l'estimateur  $\hat{m}$ . Si la fonction  $m$  est défini dans un ensemble  $\Omega$ ,  $\hat{m}_I$  sera défini comme la projection  $L_2(\Omega)$ <sup>6</sup> de  $\hat{m}$  dans l'ensemble des fonctions monotones croissantes:

$$(3.1) \quad \int_{\Omega} [\hat{m}_I(x) - \hat{m}(x)]^2 dx = \min_{g \text{ croissante}} \int_{\Omega} [g(x) - \hat{m}(x)]^2 dx.$$

La formule (3.1) implique que  $\int_0^x \hat{m}_I(t) dt$  est le plus grand convexe minorant de  $\int_0^x \hat{m}(t) dt$  et que

$$(3.2) \quad \hat{m}_I(x) = \min_{v \geq x} \max_{u \leq x} \frac{1}{v - u} \int_u^v \hat{m}(t) dt.$$

Comme  $\hat{m}_I$  est le projeté de  $\hat{m}$  sur l'ensemble des fonctions monotones croissantes, il est lui même monotone croissante. En appliquant convenablement cette opération à l'estimateur et à sa dérivée première (voir annexe), on obtiendra un nouvel estimateur qui est monotone croissante (par l'isotonisation de l'estimateur original) et convexe (par l'isotonisation de la dérivée première de l'estimateur isotonisé).

#### 3.2 Estimation par réseaux de neurones isotonisés.

On fait une isotonisation de  $m_R$  pour obtenir un nouvel estimateur  $m_{RI}$  tel que:

$$(3.3) \quad m_{RI} \left( \frac{S_t}{K} \right) = \min_{v \geq S_t/K} \max_{u \leq S_t/K} \frac{1}{v - u} \int_u^v m_R(t) dt.$$

<sup>6</sup>C'est à dire la projection suivant la norme  $L_2(\Omega)$ . On rappelle que  $\|f\|_{L_2(\Omega)} = (\int_{\Omega} (f(x))^2 dx)^{1/2}$ .

En utilisant l'expression de  $m_R$ , on obtient

$$(3.4) \quad m_{RI} \left( \frac{S_t}{K} \right) = \beta_0 + \sum_{j \in I_0} \beta_j G(\theta_j) + \min_{v \geq S_t/K} \max_{u \leq S_t/K} \frac{1}{v-u} \sum_{j \in I_1} \frac{\beta_j}{\gamma_j} \ln \left[ \frac{1 - G(\gamma_j u + \theta_j)}{1 - G(\gamma_j v + \theta_j)} \right],$$

où  $I_0$  et  $I_1$  sont des ensembles définis dans l'annexe mathématique. À partir de ceci et grâce à (4.2) de l'annexe, le ratio de couverture s'écrit:

$$(3.5) \quad \Delta_t^{RI} = \mu_R \left( \frac{S_t}{K} \right) = \min_{v \geq S_t/K} \max_{u \leq S_t/K} \frac{m_{RI}(v) - m_{RI}(u)}{v - u},$$

et

$$(3.6) \quad C_t^{RI} = K \int_0^{S_t/K} \mu_R(t) dt.$$

### 3.3 Estimation par les noyaux isotonisés.

On fait une isotonisation de l'estimateur  $m_S$ , pour obtenir un nouveau estimateur  $m_{SI}$  tel que:

$$(3.7) \quad m_{SI} \left( \frac{S_t}{K} \right) = \min_{v \geq S_t/K} \max_{u \leq S_t/K} \frac{1}{v-u} \int_u^v m_S(t) dt.$$

À partir de l'expression de  $m_S$  et grâce à (5.8) de l'annexe, on obtient le ratio de couverture:

$$(3.8) \quad \Delta_t^{SI} = \mu_S \left( \frac{S_t}{K} \right) = \min_{v \geq S_t/K} \max_{u \leq S_t/K} \frac{m_{SI}(v) - m_{SI}(u)}{v - u},$$

et le prix d'option

$$(3.9) \quad C_t^{SI} = K \int_0^{S_t/K} \mu_S(t) dt.$$

## 4 Application numérique.

### 4.1 Données simulées et observées.

Pour savoir si la méthode d'isotonisation fonctionne, on l'évalue sur des données simulées à partir de la formule de Black-Scholes. Dans ce cas, on sait que théoriquement la fonction d'évaluation d'options est monotone croissante et convexe. On peut alors évaluer si la fonction estimée sans isotonisation s'écarte des propriétés de monotonie et de convexité et jusqu'à quel point

l'isotonisation corrige ces écarts. Pour simuler des prix d'options à partir de la formule de Black-Scholes, on suppose que le prix de l'actif sous-jacent sur lequel est écrite l'option suit un mouvement brownien géométrique, soit

$$(4.1) \quad S_t = \mu S_t + \sigma S_t dW_t.$$

Le processus du calcul des termes de la série  $(S_t)_t$  est identique à celui adopté par Garcia et Gençay (1997) et Hutchinson, Lo et Poggio (1994) avec  $S_0$ , le prix initial de l'actif égal à \$50,  $\mu$  la moyenne égale à 10%, et la volatilité annuelle  $\sigma$  égale à 20%. Pour avoir des données simulées cohérentes avec les données réelles journalières, il faut utiliser une variable aléatoire normale  $(Z_i)_i$  avec une moyenne égale à  $\mu/253$  et un écart type égal à  $\sigma/\sqrt{253}$ . La série des prix  $\{S_t, t = 1, \dots, 253\}$  est calculée de la façon suivante:

$$(4.2) \quad S_t = S_0 \exp \left( \sum_{i=1}^t Z_i \right).$$

Étant donné cette trajectoire de prix d'actifs journaliers, les prix d'options ont été créés suivant les règles du Chicago Board Options Exchange (CBOE) sur l'indice S&P 500. Les options sont sur des cycles de Janvier, Février ou Mars. Le cycle de Janvier contient les mois de Janvier, Avril, Juillet et Octobre. Le cycle de Février contient les mois de Février, Mai, Août et Novembre. Le cycle de Mars contient les mois de Mars, Juin, Septembre et Décembre. Si la date d'expiration du mois courant n'est pas encore atteinte, l'option est échangée avec les dates d'expiration du mois courant, le mois suivant, et les deux mois suivants dans son cycle. Si la date d'expiration du mois courant est dépassée, l'option est échangée avec les dates d'expiration du mois suivant, et des deux mois suivants du cycle d'expiration. Pour les options, les prix d'exercice sont normalement dans des intervalles de \$2.5, \$5, ou \$10. La règle la plus utilisée dans les échanges est d'utiliser l'intervalle de \$2.5 lorsque le prix de l'actif est inférieur à \$25, l'intervalle de \$5 lorsque le prix de l'actif est compris entre \$25 et \$200 et l'intervalle de \$10 lorsque le prix de l'actif est supérieur à \$200. Lorsqu'une nouvelle date d'expiration est introduite, les deux prix d'exercice des plus proches du prix courant de l'actif sont choisis par le marché. Si l'un des deux prix est très proche du prix existant de l'actif, le troisième prix d'exercice le plus proche du prix de l'actif doit être sélectionné. Si le prix de l'actif dépasse les bornes supérieure et inférieure du prix d'exercice, un nouveau prix d'exercice va être introduit. En ce qui concerne les données observées, on utilisera les données de l'année 1988.

Dans cette section, on montrera numériquement l'efficacité de l'application de la méthode d'isotonisation surtout au niveau de la préservation des propriétés de la fonction à estimer, à savoir la monotonie croissante et la convexité. On s'intéresse à l'estimation basée sur la méthode des réseaux de neurones et l'estimation à noyaux. On considérera un ensemble de données de  $C_t$  et de  $S_t$  pour lesquelles la maturité  $\tau$  est presque constante, c'est à dire, que  $\tau/365$  soit situé dans un intervalle relativement petit. Par exemple, dans nos simulations, on prendra  $\tau/365$  compris entre 0.07 et 0.08.

Dans nos simulations, on prendra un échantillon de 311 données lorsqu'elles sont simulées et un échantillon de 229 données réelles de l'année 1988. On prendra 3 unités cachées dans les réseaux de neurones.

## 4.2 Simulations et comparaisons.

Dans les figures 1 ou 2, il y a trois graphiques qui représentent  $C_t/K$  en fonction de  $S_t/K$  en utilisant trois méthodes:

- Méthode par application de la formule de Black-Scholes,
- Méthode par application des réseaux de neurones,
- Méthode par application des réseaux de neurones et isotonisation.

On remarque que les trois graphiques sont presque confondus.

Dans les figures 3 et 4, il y a deux graphiques qui représentent le ratio de couverture en fonction de  $S_t/K$  en utilisant:

- Méthode de Balck-Scholes,
- Méthode des réseaux des neurones sans isotonisation,
- Méthode des réseaux de neurones avec isotonisation.

On remarque que la courbe qui correspond au ratio de couverture par application de la méthode des réseaux de neurones sans isotonisation n'est pas monotone croissante, par contre celle obtenue par application des réseaux de neurones et isotonisation est monotone croissante. D'où l'efficacité de l'isotonisation qui nous permet de préserver les propriétés de monotonie (figure 1 et 3) et de convexité (figure 2 et 4).

Ces remarques sont aussi valables quand on utilise l'estimation à noyaux. Les figures 5 et 6 mettent en évidence comment l'opération d'isotonisation préserve la propriété de convexité lorsqu'on estime par les noyaux.

### 4.3 Algorithme pour l'isotonisation.

Si on note  $f_I$  la régression isotonique d'une fonction  $f$ , elle est définie par:

$$(4.3) \quad \forall x \in \Omega, \quad f_I(x) = \min_{v \geq x} \max_{u \leq x} F(u, v),$$

avec

$$(4.4) \quad F(u, v) = \frac{1}{v - u} \int_u^v f(t) dt,$$

pour tout  $(u, v) \in \mathbf{R}^2$ .

Pour calculer  $f_I(x)$  pour chaque  $x$  appartenant à  $\Omega$ , on doit sélectionner des réels  $u$  et  $v$  tels que  $u \leq x$  et  $v \geq x$ . Pour cela, si on se donne un entier  $N$  suffisamment grand et un réel  $\delta$  suffisamment petit, pour chaque  $x$  de  $\Omega$ , on introduit les ensembles définis par:

$$(4.5) \quad \mathcal{U}_x = \{u_p = x - \delta p, p = 1, \dots, N\},$$

et

$$(4.6) \quad \mathcal{V}_x = \{v_q = x + \delta q, q = 1, \dots, N\}.$$

Après, pour chaque  $x$ , on construit la suite  $\{u_1^*, u_2^*, \dots, u_N^*\}$  tel que pour chaque  $q \in [1, \dots, N]$ , on a

$$(4.7) \quad F(u_q^*, v_q) = \max_{u \in \mathcal{U}_x} F(u, v_q).$$

Une fois que la suite  $(u_q^*)_q$  est construite, on va trouver  $q^*$  tel que

$$(4.8) \quad F(u_{q^*}^*, v_{q^*}) = \min_q F(u_q^*, v_q).$$

On en déduit que

$$(4.9) \quad f_I(x) = F(u_{q^*}^*, v_{q^*}).$$

Pour obtenir de bons résultats numériques, il faut que le nombre de points de discrétisation  $N$  soit très grand et le pas  $\delta$  soit très petit. Lorsqu'on prend par exemple ( $N = 100$ ) et ( $\delta = 10^{-3}$ ), on remarque que le temps de calcul est extrêmement grand. Il faut donc établir un arbitrage entre précision des résultats et temps de calcul.

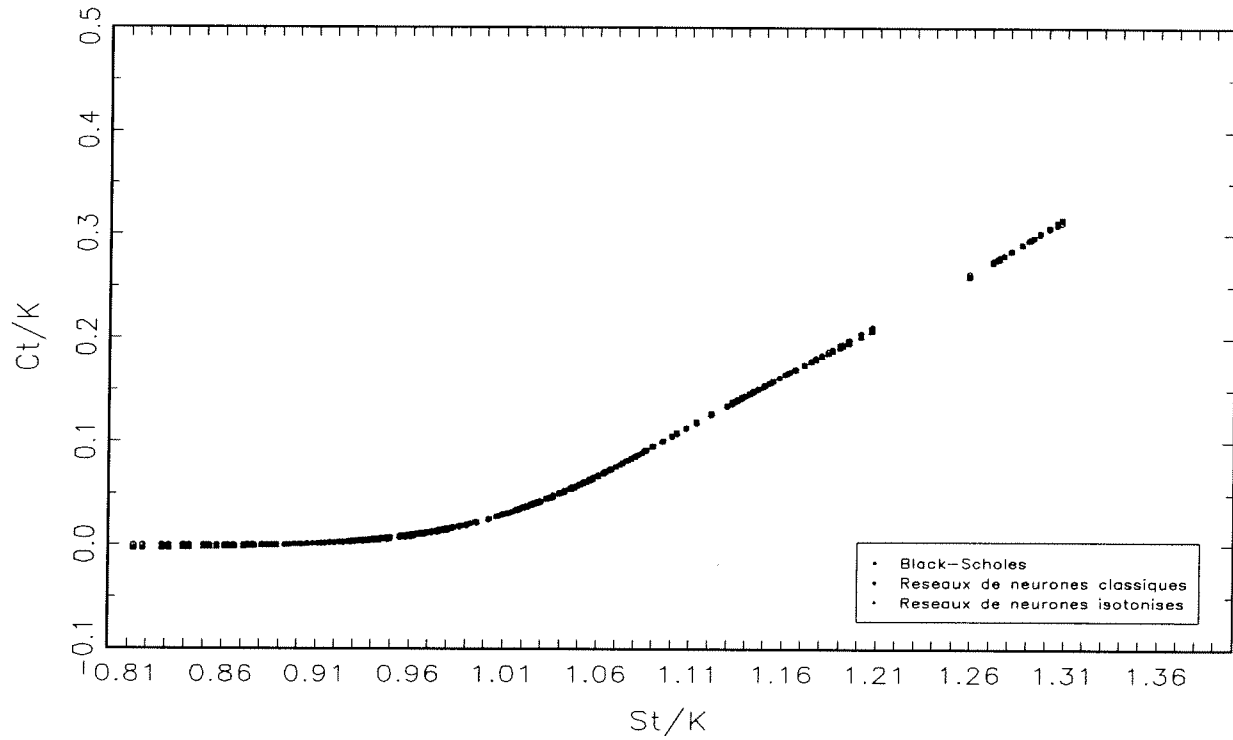
Pour sélectionner les valeurs de  $x$  telles que  $u \leq x$  et  $v \geq x$ , on ne dispose d'aucun critère qui nous permet de faire le bon choix. Pour un nombre fixe



**Figure 1**

Les courbes qui correspondent aux prix d'options en fonction du prix de l'actif sous-jacent avec trois méthodes différentes: Black-Scholes, Réseaux de neurones avec et sans isotonisation, en utilisant des données de  $S_t$  simulées par Black-Scholes et en prenant une maturité constante.

La methode des reseaux de neurones pour des donnees simules par Black-Scholes et pour une maturite constante

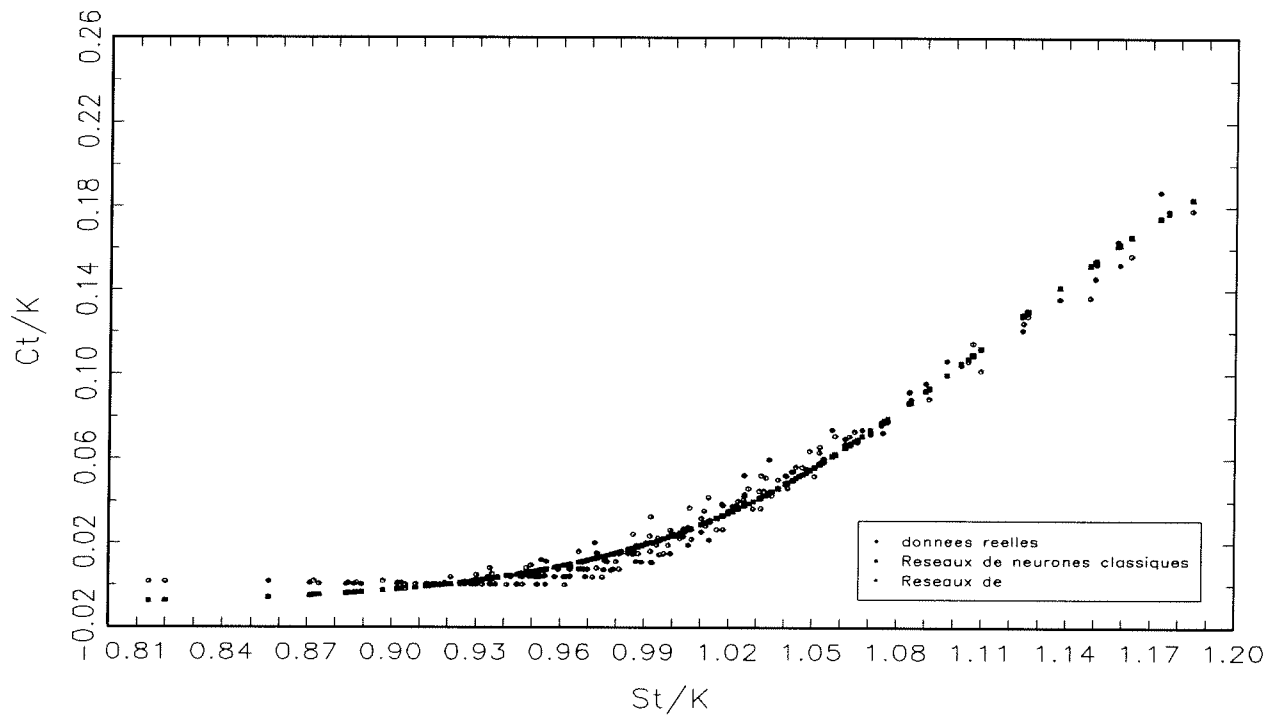


**Figure 2**

Les courbes qui correspondent aux prix d'options en fonction du prix de l'actif sous-jacent avec trois méthodes différentes: Black-Scholes, Réseaux de neurones avec et sans isotonisation, en utilisant les données réelles de  $S_t$  de l'année 1988 et en prenant une maturité constante.

GAUSS Fri Oct 23 08:34:33 1998

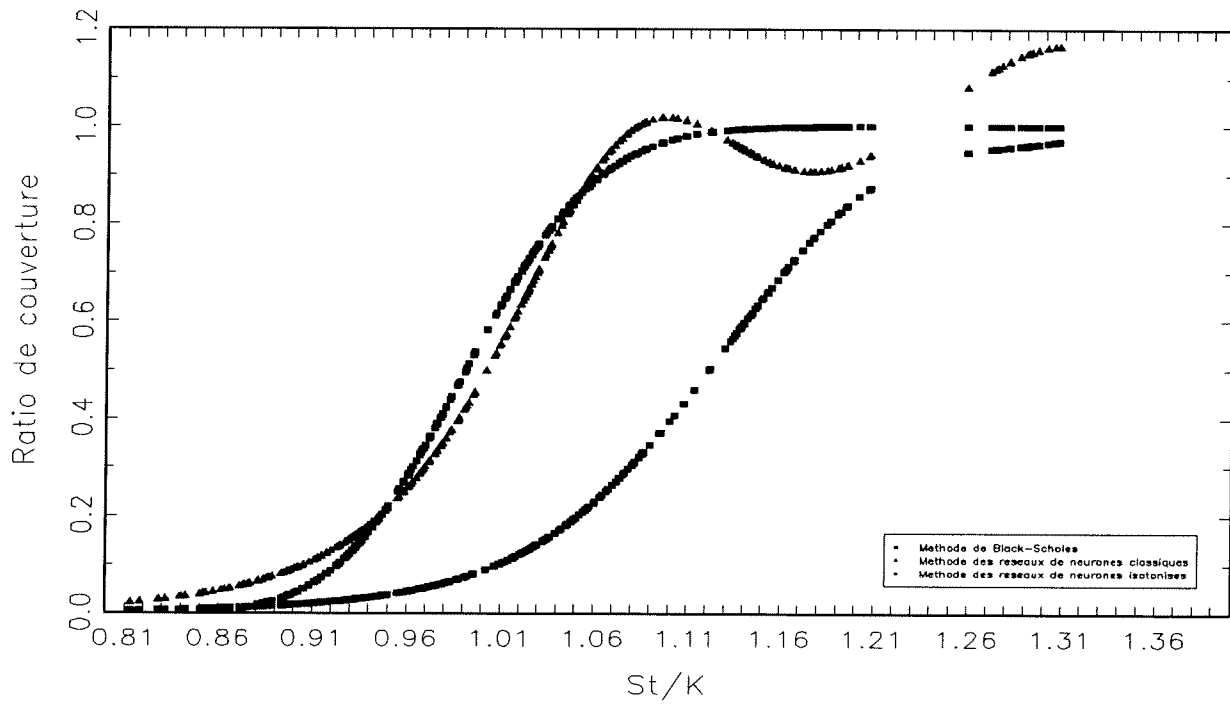
La methode des reseaux de neurones pour les donnees 88 et pour une maturite constante



**Figure 3**

Les courbes qui correspondent au ratio de couverture en fonction du prix de l'actif sous-jacent avec trois méthodes différentes:  
La méthode de Black-Scholes et les méthodes des Réseaux de neurones avec et sans l'opération d'isotonisation en utilisant des données de  $S_t$  simulées par Black-Scholes et en prenant une maturité constante.

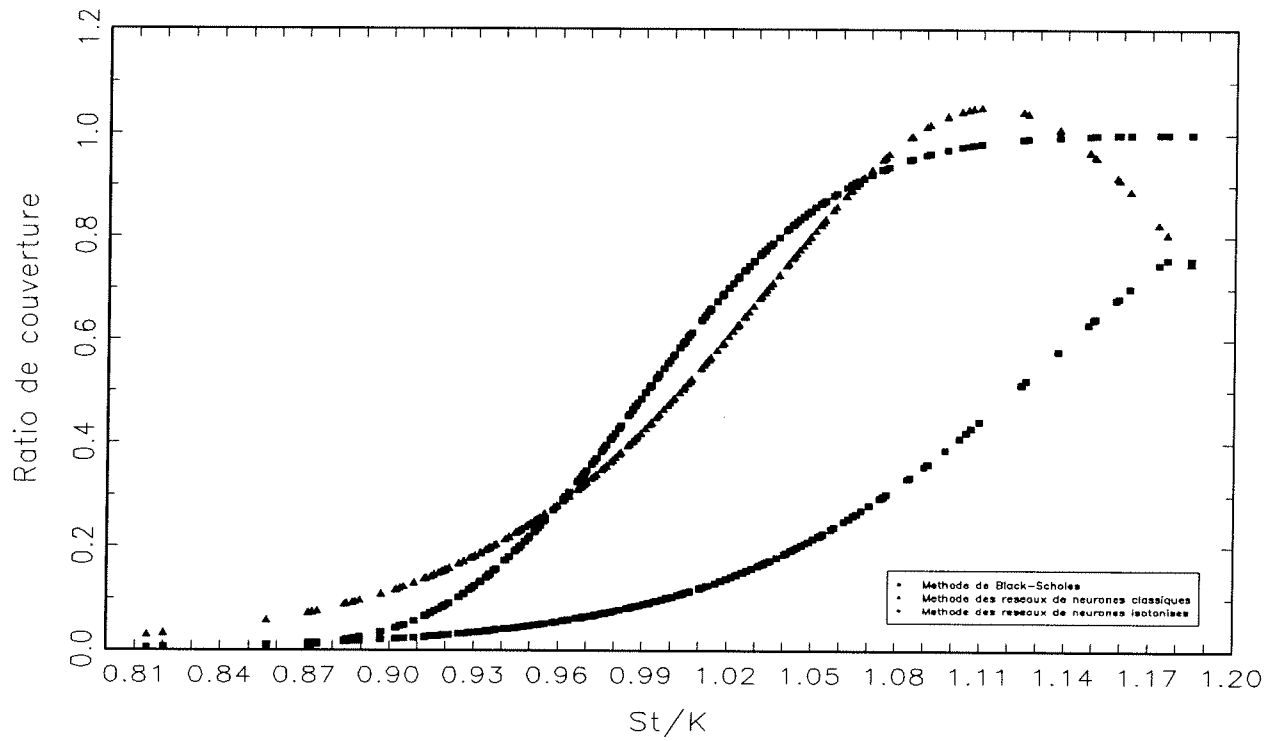
Estimation du ratio de couverture avec des données simulées en utilisant BS et les réseaux de neurones avec et sans isotonisation



**Figure 4**

Les courbes qui correspondent au ratio de couverture en fonction du prix de l'actif sous-jacent avec deux méthodes différentes:  
La méthode de Black-Scholes et les méthodes des Réseaux de neurones avec et sans l'opération d'isotonisation en utilisant les données réelles de  $S_t$  de l'année 1988 et en prenant une maturité constante.

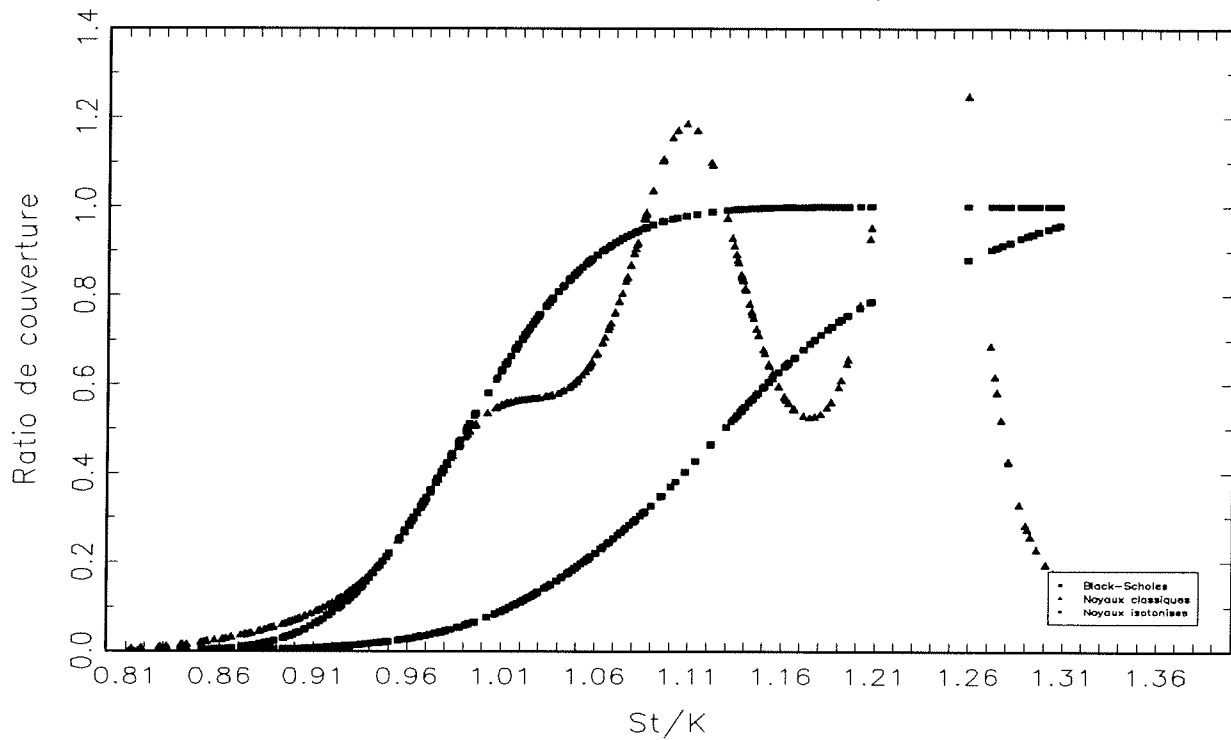
Estimation du ratio de couverture pour des données réelles en utilisant BS et les réseaux de neurones avec et sans isotonisation



**Figure 5**

Les courbes qui correspondent au ratio de couverture en fonction du prix de l'actif sous-jacent avec trois méthodes différentes:  
La méthode de Black-Scholes et les méthodes des noyaux avec et sans l'opération d'isotonisation en utilisant des données de  $S_t$  simulées par Black-Scholes et en prenant une maturité constante.

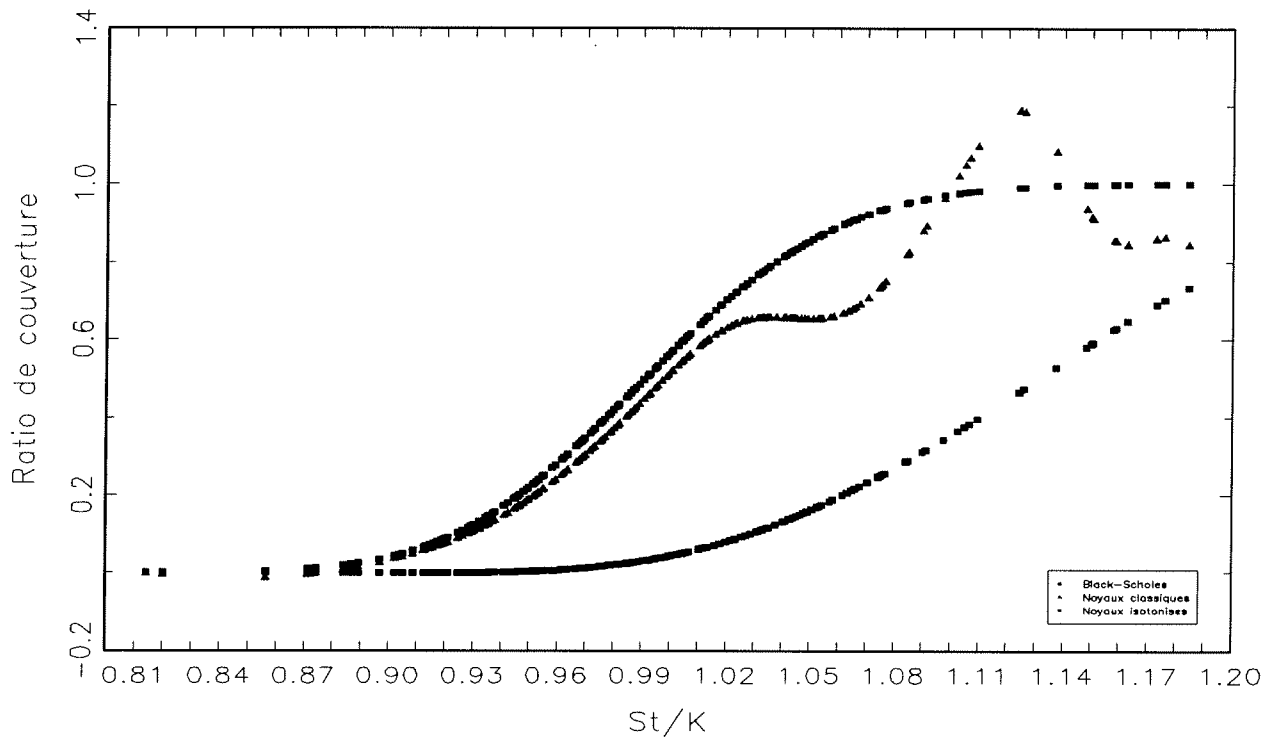
Estimation du ratio de couverture avec des données simulées en utilisant BS et la méthode des noyaux avec et sans isotonisation



**Figure 6**

Les courbes qui correspondent au ratio de couverture en fonction du prix de l'actif sous-jacent avec deux méthodes différentes:  
La méthode de Black-Scholes et les méthodes des noyaux avec et sans l'opération d'isotonisation en utilisant les données réelles de  $S_t$  de l'année 1988 et en prenant une maturité constante.

Estimation du ratio de couverture pour des données réelles en utilisant BS et les noyaux avec et sans isotonisation



des valeurs de  $x$ , on peut les prendre dans un grand intervalle mais très dispersées, ou bien dans un petit intervalle mais très concentrées. On ne peut pas dire que l'une ou l'autre des deux stratégies est la meilleure car tout dépend des données et du choix de  $N$  et du  $\delta$ .

Dans nos simulations, nous avons obtenu des résultats satisfaisants avec  $N = 20$  et  $\delta = 1/(2N)$ .

## 5 Conclusion.

Le grand avantage des résultats trouvés dans ce projet, est qu'on dispose maintenant de techniques qui nous permettent deux choses simultanément, à savoir, l'estimation et la préservation des propriétés de la fonction à estimer.

La plus grande question dans ce travail était de trouver un estimateur de la fonction d'évaluation d'options qui soit convexe en partant de deux types d'estimation de base, à savoir l'apprentissage par les réseaux de neurones et les estimateurs à noyaux. Les techniques d'isotonisation qui ont été utilisé dans ce rapport nous ont permis de répondre à cette question qui s'avère très importante dans l'estimation du ratio de couverture. D'après les figures des ratios de couverture, on remarque que la méthode de Black-Scholes répond à cette question, mais notre but était de trouver une méthode non-paramétrique qui répond à cette même question.

Ainsi, on estime mieux le ratio de couverture, ce qui permet de diminuer l'erreur de couverture dans une stratégie dynamique de couverture.



## Annexe Mathématique

### 1. Introduction:

On se propose d'estimer une fonction  $m$  qui intervient dans un modèle non-paramétrique de la forme:

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \dots, N$$

Ce type de modèle a été étudié par plusieurs statisticiens sous différentes hypothèses de régularité et de monotonie sur la fonction  $m$ . On peut citer par exemple le papier de Mammen (1991), où pour  $m$  suffisamment régulière et monotone croissante, il propose un estimateur basé sur l'estimation par les noyaux pour préserver la régularité, et sur une opération d'isotonisation pour préserver la monotonie. Il a montré qu'en alternant ces deux opérations, il trouve un nouvel estimateur qui est asymptotiquement équivalent au premier avec une précision d'ordre un. On peut citer aussi Mukerjee (1988) qui propose une procédure hybride avec laquelle il obtient un estimateur monotone croissante qui a les mêmes propriétés que la fonction estimée intervenant dans la régression non paramétrique. Wright (1982) modifie l'estimation par maximum de vraisemblance d'une régression avec erreurs normalement distribuées en groupant des observations adjacentes et en isotonisant leurs moyennes. Par ailleurs Stone (1982) étudie le taux de convergence pour une régression non paramétrique. Finalement, Silverman (1985) a appliqué une régularisation par Splines à ce type de modèles.

Dans ce présent travail, on supposera que la fonction  $m$  est suffisamment régulière, monotone croissante et convexe. On adoptera la même approche que Mammen (1991) et on l'appliquera à des estimateurs de  $m$  et de  $m'$ .

En ce qui concerne les réseaux de neurones, plusieurs auteurs se sont intéressés à leurs propriétés théoriques lorsque la fonction d'activation est du type sigmoïdale<sup>7</sup>. Par exemple, Stinchcombe et White (1989) établissent rigoureusement que les réseaux de neurones sont capables d'approximer toute fonction mesurable au sens de Borel, définie sur ensemble dans un espace de dimension finie vers un autre avec n'importe quelle degré de précision désirée. On déduit à partir d'ici que les réseaux de neurones présentent une classe des approximateurs universels. Hornik, Stinchcombe et White (1990) donnent les conditions sous lesquelles les réseaux de neurones sont capables d'estimer n'importe quelle fonction et ses dérivées. Il faut remarquer qu'ils ont montré que ceci reste vrai même pour les fonctions qui ne

---

<sup>7</sup> $G$  est une fonction sigmoïdale si elle est définie de  $\mathbf{R}$  vers  $[0, 1]$  telle que  $G(a)$  tend vers 0 lorsque  $a$  tend vers  $-\infty$ ,  $G(a)$  tend vers 1 lorsque  $a$  tend vers  $+\infty$  et  $G$  est monotone croissante.

sont pas différentiables au sens fort mais qui possèdent des dérivées au sens faible, par exemple les fonctions différentiables par morceaux. Il existe aussi des résultats théoriques même lorsque la fonction d'activation n'est pas du type sigmoïdale, mais juste une fonction non constante, bornée et parfois continue. Par exemple, Hornik (1990) montre que ces types de réseaux de neurones sont des estimateurs universels par rapport au critère de la performance  $L^p$ . Il donne aussi des conditions sous lesquelles cette classe d'estimateurs peut approximer une fonction continue et ses dérivées qui sont identiques à celles citées auparavant. Gallant et White (1991) étudient cette classe d'estimateurs dans des espaces de Sobolev et montrent plusieurs propriétés de régularité. Toutefois, pour que les réseaux de neurones puisse approcher convenablement les dérivées d'une fonction inconnue, il est nécessaire que le nombre d'unités neuronales augmente avec le nombre d'observations. L'application de ce critère en échantillon fini pourrait conduire à un surajustement de la fonction aux données, ce qui conduirait à la perte de toute capacité à bien prévoir avec une telle fonction. Il est donc important de pouvoir préserver, avec un nombre limité d'unités neuronales, les propriétés recherchées de la fonction et de ses dérivées.

En ce qui concerne les estimateurs à noyaux, leur plus grand avantage est qu'ils sont dotés de plusieurs résultats théoriques accessibles ayant un rapport avec leur régularité et leur convergence asymptotique, mais par contre leur utilisation devient très compliquée numériquement si le nombre de variables devient grand.

Sachant que les propriétés de monotonie et de convexité de  $m$  sont très importantes dans plusieurs applications en particulier dans l'évaluation du prix d'options et du ratio de couverture en fonction du prix de l'actif sous-jacent, on doit trouver des estimateurs de  $m$  qui préservent ces propriétés. Pour cela, on se basera sur les réseaux de neurones et sur les estimateurs à noyaux et on appliquera l'opération d'isotonisation aux estimateurs de  $m$  pour préserver la monotonie et aux estimateurs de  $m'$  pour préserver la convexité.

## 2. Énoncé du problème.

Désignons par  $(x_i)_i$  et  $(y_i)_i$  des observations qui correspondent respectivement aux inputs et outputs, et par  $(\varepsilon_i)_i$  des variables aléatoires supposées indépendantes et identiquement distribuées telle que  $E(\varepsilon_i) = 0$ . Supposons que les inputs sont dans un intervalle fermé  $\Omega$  de  $\mathbf{R}$ . Considérons le modèle suivant:

$$(2.1) \quad y_i = m(x_i) + \varepsilon_i, \quad i = 1, \dots, N$$

où  $m$  est une fonction définie sur  $\Omega$  à valeurs dans  $\mathbf{R}$ , inconnue, qui vérifie

les hypothèses suivantes:

$$(2.2) \quad m \text{ est deux fois continuellement différentiable : } m \in \mathcal{C}^2(\Omega),$$

$$(2.3) \quad m \text{ est croissante : } m'(x) \geq 0 \quad \forall x \in \Omega,$$

$$(2.4) \quad m \text{ est convexe : } m''(x) \geq 0 \quad \forall x \in \Omega,$$

$$(2.5) \quad m(0) = 0.$$

### 3. Rappels de quelques propriétés d'isotonisation.

Les régressions isotoniques ont été introduites par R. E. Barlow, D. J. Bartholomew, J. M. Bremner et H. D. Brunk, qui ont proposé la théorie et l'application des régressions isotoniques en 1972. Le problème de la régression isotonique et de son dual ont été étudiés par R. E. Barlow et H. D. Brunk (1972).

On va donner quelques notions de cette théorie pour comprendre son utilisation. On commence par donner la définition d'une fonction isotonique dans les cas discret et continu:

**Définition 1 (cas discret)** Soit  $X$  un ensemble fini  $\{x_1, \dots, x_k\}$  et soit  $\prec$  un ordre simple sur cet ensemble tel que  $x_1 \prec x_2 \prec \dots \prec x_k$ . La fonction réelle  $f$  sur  $X$  est isotonique si pour tout  $x, y \in X$  tels que  $x \prec y$  on a  $f(x) \leq f(y)$ .

**Définition 2 (cas continu)** Soit  $X$  un intervalle de  $\mathbf{R}$ , la fonction réelle  $f$  est isotonique sur  $X$  est équivalent à  $f$  est monotone croissante sur  $X$ .

Maintenant, on va donner la définition d'une régression isotonique dans les cas discret et continu:

**Définition 3 (cas discret)** Soit  $g$  une fonction sur  $X$  et  $\omega$  une fonction positive sur  $X$ . Une fonction isotonique  $g^*$  sur  $X$  est une régression isotonique de  $g$  avec les poids  $\omega$  par rapport à l'ordre simple  $x_1 \prec x_2 \prec \dots \prec x_k$ , si elle minimise la somme:

$$(3.1) \quad \sum_{x \in X} [g(x) - f(x)]^2 \omega(x),$$

lorsque  $f$  se trouve dans la classe des fonctions isotoniques.

**Définition 4 (cas continu)** Une fonction isotonique  $g^*$  sur un ensemble  $X$  est une régression isotonique de  $g$  si elle réalise le minimum de la quantité:

$$(3.2) \quad \int_X [g(x) - f(x)]^2 dx$$

dans l'ensemble des fonctions isotoniques  $f$  sur  $X$ .

Les deux théorèmes suivants donnent des propriétés sur les régressions isotoniques dans les cas discret et continu:

**Théorème 1 (cas discret)** Soit  $X$  un ensemble fini et soient  $g$  et  $\omega$  des fonctions données sur  $X$ , avec  $\omega > 0$ . Soit  $\mathcal{C}$  un cône convexe des fonctions sur  $X$ . Alors la fonction  $u$  dans  $\mathcal{C}$  minimise

$$(3.3) \quad \sum_x [g(x) - f(x)]^2 \omega(x),$$

pour  $f$  dans  $\mathcal{C}$  si et seulement si

$$(3.4) \quad \sum_x [g(x) - u(x)] u(x) \omega(x) = 0,$$

et

$$(3.5) \quad \sum_x [g(x) - u(x)] f(x) \omega(x) \leq 0,$$

pour tout  $f \in \mathcal{C}$ . Si  $u$  satisfait ces conditions alors, on a

$$(3.6) \quad \sum_x [g(x) - f(x)]^2 \omega(x) \geq \sum_x [g(x) - u(x)]^2 \omega(x) + \sum_x [u(x) - f(x)]^2 \omega(x),$$

pour tout  $f \in \mathcal{C}$ .

**Théorème 2 (cas continu)** Soit  $\mu$  une fonction inconnue sur un ensemble fini  $X$ , supposée isotonique par rapport au quasi-ordre sur  $X$ . Soit  $g$  un estimateur de  $\mu$  et  $g^*$  la régression isotonique de  $g$ , alors:

$$(3.7) \quad \int_X [\mu(x) - g^*(x)]^2 dx \leq \int_X [\mu(x) - g(x)]^2 dx.$$

**Remarque 3.1.** La régression isotonique d'un estimateur d'une fonction monotone croissante  $\mu$  est un estimateur monotone croissante de cette fonction, qui est plus précis que l'estimateur qui a subi la régression.

**Remarque 3.2.** Pour tout estimateur d'une fonction isotonique, il existe une et une seule régression isotonique.

#### 4. Estimation en utilisant les réseaux de neurones et un processus d'isotonisation:

En ce qui concerne les réseaux de neurones, plusieurs auteurs se sont intéressés à leurs propriétés théoriques lorsque la fonction d'activation est du type sigmoïdale.

On note  $\hat{m}_R$  l'estimateur de la fonction  $m$  obtenu à partir des estimateurs des réseaux de neurones et par un processus d'isotonisation. La construction se fait en quatre étapes:

##### 4.1. Première étape: estimation par réseaux de neurones ordinaires.

On note  $m_R$  l'estimateur de  $m$  du modèle (2.1) par les réseaux de neurones et  $G$  la fonction d'activation du type sigmoïdal et choisie de la forme:

$$(4.1) \quad G(x) = \frac{1}{1 + e^{-x}}.$$

L'estimateur  $m_R$  s'exprime de la façon suivante:

$$(4.2) \quad \forall x \in \Omega, \quad m_R(x) = \beta_0 + \sum_{j=1}^d \beta_j G(\gamma_j x + \theta_j),$$

où  $\beta_0, \beta_j, \theta_j$  et  $\gamma_j$ , avec  $j = 1, \dots, d$  sont les solutions du problème de minimisation suivant:

$$(4.3) \quad \min_{\tilde{\beta}_0, \tilde{\beta}_j, \tilde{\gamma}_j, \tilde{\theta}_j} \sum_{i=1}^N \left[ y_i - \tilde{\beta}_0 - \sum_{j=1}^d \tilde{\beta}_j G(\tilde{\gamma}_j x_i + \tilde{\theta}_j) \right]^2,$$

et  $d$  est le nombre d'unités neuronales.

##### 4.2. Deuxième étape: L'isotonisation de l'estimateur $m_R$ .

On note  $m_{RI}$  la régression isotonique de l'estimateur  $m_R$ . Elle est définie comme la projection  $L_2(\Omega)$  de  $m_R$  sur la classe des fonctions monotones croissantes, plus précisément:

$$(4.4) \quad \int_{\Omega} [m_{RI}(x) - m_R(x)]^2 dx = \min_{g \text{ croissante}} \int_{\Omega} [g(x) - m_R(x)]^2 dx.$$

Si on suppose que  $\Omega = [0, \bar{\omega}]$ , où  $\bar{\omega}$  est un réel positif, la relation (4.4) implique que  $\int_0^x m_{RI}(t) dt$  est le plus grand minorant convexe de  $\int_0^x m_R(t) dt$  et que:

$$(4.5) \quad m_{RI}(x) = \min_{v \geq x} \max_{u \leq x} \frac{1}{v - u} \int_u^v m_R(t) dt.$$

Sachant l'expression de  $m_R$  donnée par (4.2), on trouve que:

$$(4.6) \quad \int_u^v m_R(t)dt = \beta_0(v - u) + \sum_{j=1}^d \beta_j \int_u^v G(\gamma_j t + \theta_j)dt.$$

Posons:

$$(4.7) \quad I_0 = \{j = 1, \dots, d \mid \gamma_j = 0\},$$

et

$$(4.8) \quad I_1 = \{j = 1, \dots, d \mid \gamma_j \neq 0\}.$$

En faisant un changement de variables en posant  $z = \gamma_j t + \theta_j$  et sachant que  $G'(x) = G(x)(1 - G(x))$ , on trouve que lorsque  $j \in I_1$ , on a

$$(4.9) \quad \int_u^v G(\gamma_j t + \theta_j)dt = \frac{1}{\gamma_j} \ln \left[ \frac{1 - G(\gamma_j u + \theta_j)}{1 - G(\gamma_j v + \theta_j)} \right],$$

et lorsque  $j \in I_0$ , on a

$$(4.10) \quad \int_u^v G(\gamma_j t + \theta_j)dt = G(\theta_j)(v - u).$$

Les relations (4.5), (4.9) et (4.10) impliquent que:

$$m_{RI}(x) = \beta_0 + \sum_{j \in I_0} \beta_j G(\theta_j) + \min_{v \geq x} \max_{u \leq x} \frac{1}{v - u} \sum_{j \in I_1} \frac{\beta_j}{\gamma_j} \ln \left[ \frac{1 - G(\gamma_j u + \theta_j)}{1 - G(\gamma_j v + \theta_j)} \right].$$

**Remarque 4.1.** On ne peut s'arrêter ici, car même si  $m_{RI}$  est un estimateur de  $m$  qui est plus précis que  $m_R$  en vertu des propriétés sur l'isotonisation et de plus, il est monotone croissant, il n'y a rien qui nous garantie qu'il est convexe, pour cela, on doit faire une isotonisation de l'estimateur de  $m$  (car  $m' > 0$ ) et de celui de  $m'$  car ( $m'' > 0$ ).

### 4.3. Troisième étape: L'isotonisation de la dérivée de $m_{RI}$ .

Sous réserve que  $m'_{RI}$  estime  $m'$ , (ceci est fort probable car d'une part  $m'_R$  estime  $m'$  sous certaines conditions à savoir qu'il faut que le nombre des unités neuronales augmente asymptotiquement avec le nombre d'observations, et d'autre part,  $m_{RI}$  est plus précis que  $m_R$ ), on peut faire alors de nouveau une régression isotonique de  $m'_{RI}$  qu'on notera  $\mu_R$ , et ceci parce que pour tout  $x$  dans  $\Omega$ ,  $m''(x) \geq 0$ . Donc  $\mu_R$  est définie par

$$(4.12) \quad \mu_R(x) = \min_{v \geq x} \max_{u \leq x} \frac{1}{v - u} \int_u^v m'_{RI}(t)dt,$$

ce qui implique que

$$(4.13) \quad \mu_R(x) = \min_{v \geq x} \max_{u \leq x} \frac{m_{RI}(v) - m_{RI}(u)}{v - u}.$$

#### 4.4. Quatrième étape : Définition de l'estimateur $\hat{m}_R$ .

Sachant que  $m(0) = 0$ , on définit la fonction  $\hat{m}$  par:

$$(4.14) \quad \forall x \in \Omega, \quad \hat{m}_R(x) = \int_0^x \mu_R(t) dt.$$

Pour calculer  $\hat{m}_R$ , on fait une intégration numérique en utilisant la formule de Simpson qui s'avère plus précise que plusieurs autres formules d'intégration. Pour cela, pour chaque  $x$ , on fait une subdivision de l'intervalle  $[0, x]$  en posant:

$$\forall x \in \Omega, \quad \forall j = 1, \dots, N + 1 \quad x_j = \frac{x(j-1)}{N}.$$

on obtient alors que pour tout  $x \in \Omega$ ,

$$(4.15) \quad \hat{m}_R(x) \approx \sum_{j=1}^N \frac{\mu_R(x_j) + 4\mu_R((x_j + x_{j+1})/2) + \mu_R(x_{j+1})}{6} (x_{j+1} - x_j).$$

#### 4.5. Propriétés de l'estimateur $\hat{m}_R$ .

Dans cette partie, on montrera que l'estimateur  $\hat{m}_R$  possède les mêmes propriétés que la fonction  $m$ . Et ceci revient aux propriétés de l'isotonisation, à savoir que la régression isotonique d'un estimateur d'une fonction isotonique est une fonction isotonique. C'est cette propriété qui fait la force de notre estimateur, car on ne veut pas seulement estimer la fonction  $m$ , mais on veut que l'estimateur préserve ses propriétés.

Premièrement, sachant que  $m_{RI}$  est la régression isotonique de  $m_R$ , donc elle est croissante, par conséquent:

$$(4.16) \quad \forall v > u, \quad \frac{m_{RI}(v) - m_{RI}(u)}{v - u} > 0,$$

et en regardant la définition de  $\mu_R$  donnée par (4.13), on déduit que  $\mu_R$  est positive. D'autre part, grâce à la définition (4.14) de  $\hat{m}_R$ ,  $\mu_R$  est la dérivée de  $\hat{m}_R$ , par conséquent  $\hat{m}_R$  est monotone croissante.

Deuxièmement,  $\mu_R$  est monotone croissante (car elle est la régression isotonique de  $m'_{RI}$ ), donc  $\hat{m}'_R$  est monotone croissante, par conséquent  $\hat{m}_R$  est convexe.

Troisièmement, d'après la définition (4.14) de  $\hat{m}_R$ , on a  $\hat{m}_R(0) = 0$ .

Ceci montre que notre estimateur préserve les propriétés de monotonie, de convexité et de condition initiale qui s'avèrent très utiles dans l'application qu'on présentera par la suite.

### 5. Estimation en utilisant les noyaux et un processus d'isotonisation.

En ce qui concerne les estimateurs à noyaux, leur plus grand avantage est qu'ils sont dotés de plusieurs résultats théoriques accessibles ayant un rapport avec leur régularité et leur convergence asymptotique, mais par contre leur utilisation devient très compliquée numériquement si le nombre de variables devient grand. On ne peut pas dire si les propriétés de la fonction sont conservées si on utilise ce genre d'estimateurs, d'où encore une fois l'intérêt d'isotoniser les estimateurs de  $m$  et de  $m'$  en utilisant les noyaux.

On note  $\hat{m}_S$  l'estimateur de la fonction  $m$  obtenu en utilisant les noyaux et un processus d'isotonisation. La construction de cet estimateur se fait en quatre étapes:

#### 5.1. Première étape: estimation par les noyaux.

On note  $m_S$  l'estimateur de  $m$  du modèle (2.1) obtenu par les noyaux classiques, il s'écrit:

$$(5.1) \quad m_S(x) = \frac{1}{N} \sum_{i=1}^N \omega_{i,N}(x) y_i,$$

où

$$(5.2) \quad \omega_{i,N}(x) = K_h(x - x_i) / g_h(x),$$

et

$$(5.3) \quad g_h(x) = \frac{1}{N} \sum_{i=1}^N K_h(x - x_i).$$

Le réel  $h$  est un paramètre de régularisation. Si  $h$  est petit, la moyenne sera faite par rapport à des voisinages étroits autour des  $x_i$ . Si  $h$  est grand, la moyenne sera faite par rapport à des voisinages larges autour des  $x_i$ . Toutefois, on peut contrôler la somme des moyennes en ajustant le paramètre  $h$ .

En substituant les relations (5.2) et (5.3) dans (5.1), on trouve que

$$(5.4) \quad m_S(x) = m_S^h(x) = \frac{\sum_{i=1}^N K_h(x - x_i) y_i}{\sum_{i=1}^N K_h(x - x_i)}.$$



Dans la plupart des cas, le choix du noyau est le noyau Gaussien, soit

$$(5.5) \quad K_h(x) = \frac{1}{h\sqrt{2\pi}} e^{-\frac{x^2}{2h^2}}.$$

### 5.2. Deuxième étape: L'isotonisation de $m_S$ .

L'isotonisation de l'estimateur à noyaux de  $m$  a été faite par Mammen (1991). Donc, si  $\Omega$  désigne l'intervalle fermé de  $\mathbf{R}$  contenant tous les  $x_i$  avec  $i = 1, \dots, N$ , et  $m_{SI}$  le noyau isotonisé, alors il est défini comme la projection  $L_2(\Omega)$  de  $m_S$  sur la classe des fonctions monotones croissantes, soit

$$(5.6) \quad \int_{\Omega} [m_{SI}(x) - m_S(x)]^2 dx = \min_{g \text{ croissante}} \int_{\Omega} [g(x) - m_S(x)]^2 dx.$$

#### Remarque 5.1.

L'estimateur  $m_{SI}$  est légèrement différent de l'estimateur introduit par Friedman et Tibshirani (1984). Un autre estimateur similaire a été proposé par Wright (1982) [ voir aussi Barlow et Van Zwet (1969, 1970)].

Si on suppose que  $\Omega = [0, \bar{\omega}]$ , où  $\bar{\omega}$  est un réel positif, la relation (5.6) implique que  $\int_0^x m_{SI}(t) dt$  est le plus grand minorant convexe de  $\int_0^x m_S(t) dt$  et que:

$$(5.7) \quad m_{SI}(x) = \min_{v \geq x} \max_{u \leq x} \frac{1}{v-u} \int_u^v m_S(t) dt.$$

En ce qui concerne le paramètre de régularisation, puisque la fonction  $m$  est supposée deux fois continuellement différentiable, le choix optimal de la fenêtre  $h$  qui intervient dans l'estimateur à noyaux est de l'ordre  $N^{-1/5}$ . Donc, sans perte de généralité, on prendra dans les simulations

$$h = N^{-1/5}.$$

#### Remarque 5.2.

E. Mammen (1991) a proposé un autre estimateur  $m_{IS}$  en interchangeant les étapes de régularisation et d'isotonisation, et il a montré que les deux estimateurs sont asymptotiquement équivalents avec une erreur d'ordre 1. De plus, l'estimateur  $m_{IS}$  a une erreur quadratique moyenne plus petite que celle de  $m_{SI}$  si et seulement si la fonction à noyaux utilisée dans l'estimation n'est pas suffisamment régulière.

### 5.3. Troisième étape: L'isotonisation de la dérivée de $m_{SI}$ .

Sous réserve que  $m'_{SI}$  estime  $m'$ , on peut alors faire de nouveau une nouvelle régression isotonique de  $m'_{SI}$  qu'on va noter  $\mu_S$ , et ceci c'est parce que  $m'$  est monotone croissante puisque  $m$  est convexe. Donc  $\mu_S$  est définie par

$$(5.8) \quad \mu_S(x) = \min_{v \geq x} \max_{u \leq x} \frac{1}{v-u} \int_u^v m'_{SI}(t) dt,$$

ce qui implique que

$$(5.9) \quad \mu_S(x) = \min_{v \geq x} \max_{u \leq x} \frac{m_{SI}(v) - m_{SI}(u)}{v-u}.$$

#### 5.4. Quatrième étape : Définition de l'estimateur $\hat{m}_S$ .

Sachant que  $m(0) = 0$ , on définit la fonction  $\hat{m}_S$  par:

$$(5.10) \quad \forall x \in \Omega, \quad \hat{m}_S(x) = \int_0^x \mu_S(t) dt.$$

Pour calculer  $\hat{m}_S$ , on fait une intégration numérique en utilisant la formule de Simpson. Pour cela, pour chaque  $x$ , on fait une subdivision de l'intervalle  $[0, x]$  en posant:

$$\forall x \in \Omega, \quad \forall j = 1, \dots, N+1 \quad x_j = \frac{x(j-1)}{N}.$$

on obtient que pour tout  $x \in \Omega$ ,

$$(5.11) \quad \hat{m}_S(x) \approx \sum_{j=1}^N \frac{\mu_S(x_j) + 4\mu_S((x_j + x_{j+1})/2) + \mu_S(x_{j+1})}{6} (x_{j+1} - x_j).$$

#### 5.5. Propriétés de l'estimateur $\hat{m}_S$ .

Dans cette partie, on montrera que l'estimateur  $\hat{m}_S$  possède les mêmes propriétés que la fonction  $m$ .

Premièrement, sachant que  $m_{SI}$  est la régression isotonique de  $m_S$ , donc elle est croissante, par conséquent:

$$(5.12) \quad \forall v > u, \quad \frac{m_{SI}(v) - m_{SI}(u)}{v-u} > 0,$$

et en regardant la définition de  $\mu_S$  donnée par (5.9), on déduit que  $\mu_S$  est positive. D'autre part, grâce à la définition (4.14) de  $\hat{m}_S$ ,  $\mu_S$  est la dérivée de  $\hat{m}_S$ , par conséquent  $\hat{m}_S$  est monotone croissante.

## References

- [1] Ait-Sahalia, Y. and A. Lo (1996), "Nonparametric estimation of State-Price Densities Implicit in Financial Asset Prices", Working Paper LFE-1015-96, MIT Laboratory for Financial Engineering.
- [2] Barlow, R. E. Bartholomew, D. J., Bremner, J. M. and Brunk, H. D.,(1972). "Statistical Inference Under Order Restrictions". Wiley, New York.
- [3] Barlow, R. E. and Van Zwet, W. R.(1969). "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part II: Asymptotic distributions". Operations Research Center Report ORC 69-10, Univ. California.
- [4] Barlow, R. E. and Van Zwet, W. R.(1970). "Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: Strong consistency". In Nonparametric Techniques in Statistical Inference (M. L. Puri, ed.) 159-173. Cambridge Univ. Press
- [5] Barlow, R. E., and V. A. Ubhaya(1971). "Isotonic Approximation". In J. S. Rustagi (Ed.), Optimizing Methods in Statistics, Academic Press, New York.
- [6] Barlow, R. E., and H. D. Brunk(1972). "The isotonic regression problem and its dual". J.Am.statist.Ass., 67, 140-147.
- [7] Cheng, K. F. and Lin, P. E. (1981). "Nonparametric estimation of a regression function." Z. Wahrsch. Verw. Gebiete, 57, 223-233.
- [8] Friedman, J. and Tibshirani, R.(1984). "The monotone smoothing of scatter plots". Technometrics, 26, 243-350.
- [9] Gallant, A. R., and White, H.(1988). "There exists a neural network that does not make avoidable mistake". In IEEE Second International Conference on Neural Network, 657-664.
- [10] Gallant, A. R., and White, H.(1992). "On learning the derivatives of an unknown mapping with multilayer feedforward networks", Neural Networks, 5, 129-138.
- [11] Garcia, R. and E. Renault(1995), "Risk Aversion, Intertemporal Substitution and Option Pricing", Working paper, Université de Montréal.
- [12] Garcia, R. and R. Gençay(1997), "Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint", Working paper, Université de Montréal.
- [13] Gaylord, CK. and Ramirez, D.E. (1991). "Monotone regression splines for smoothed bootstrapping, *Computational Statistics Quarterly* 6(2):85-97.

- [14] Hecht-Nielsen, R. (1989). "Theory of the back propagation neural network". In proceedings of International Joint Conference on Neural Network, 593-606. San Diego:SOS Printing.
- [15] Hildenbrand, K. and Hildenbrand, W. (1985), " On the mean income effect: A data analysis of the U.K. family expenditure survey". In contributions to Mathematical Economics.(W. Hildenbrand and A. Mascollel, eds.) 247-268. North-Holland, Amsterdam.
- [16] Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer feed-forward networks are universal approximators", *Neural Networks*, 2, 359-366.
- [17] Hornik, K., Stinchcombe, M., and White, H. (1990). "Universal approximation of an unknown mapping and its derivatives using multilayer feedforworks". *Neural Networks*, 3, 551-560.
- [18] Hornik, K.,(1991), "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, 4, 251-257.
- [19] Hutchinson, J. M., A. W. Lo and T. Poggio (1994), "A Nonparametric approach to pricing and hedging derivative securities via learning network", *Journal of Finance*, 3, 851-889.
- [20] Irie, B., and Miyake, S.(1988). "Capabilities of three layer perceptrons". In IEEE Second Internatinal Conference on Neural Networks, 641-648.
- [21] Leurgans, S. (1982). "Asymptotic distributions of slope-of-greatest-convex-minorant estimators". *Ann. Statist.* 10, 287-296.
- [22] Leurgans, S. (1986). "Isotonic M-estimation". *Advances in Order Restricted. Inference. Lecture Notes in Statist.* 37, 48-68. Springer, New York.
- [23] Mammen E., (1991). "Estimating A Smooth Monotone Regression Function". *Ann. Statist.* Vol 19, No 2, 724-740.
- [24] Mammen E., (1991). "Nonparametric Regression Under Qualitative Smoothness Assymptions". *Ann. Statist.* Vol 19, No 2, 741-759.
- [25] Mammen E.and Thomas-Agnan, C. (1996). "Smoothing splines and shape restrictions", *Discussion Paper 87*, Sonderforschungsbereich 373, Humboldt-Universitat zu Berlin. **URL:** ftp://amadeus.wiwi.hu-berlin.de/pub/papers/sfb373/sfb1996/dpsfb960087.ps.Z
- [26] Matzkin, R.L. (1991). "Semiparametric estimation of monotone and concave utility functions for polychotomous choice models", *Econometrica* 59(5):1315-1327.

- [27] Matzkin, R.L. (1994). "Restrictions of economic theory in nonparametric methods", in R.F. Engle and D.L. McFadden (eds), *Handbook of Econometrics*, Vol. 4, North Holland, Amsterdam, chapter 42, pp.2523-2558.
- [28] Maxwell, T., Giles, G. L., Lee, Y. C., and Chen, H. H.(1986). "Nonlinear dynamics of artificial neural systems". In J. Denker (Ed.), *Neural networks for computing*. New York: American Institute of Physics
- [29] Miller, D. R. and Sofer, A. (1986). "Least-squares regression under convexity and higher-order difference constraints with application to software reliability". *Advances in Order Restricted Inference. Lecture Notes in Statist.* 37, 91-124, Springer, New York.
- [30] Mukerjee, H. (1988). "Monotone nonparametric regression". *Ann. Statist.* 16 741-750.
- [31] Müller, H.-G.(1988). "Nonparametric Regression Analysis of Longitudinal Data". *Lecture Notes in Statist.* 46. Springer. New York.
- [32] Müller, H.-G. and U. Stadtmüller (1987) "*Estimation of Heteroscedasticity in regression Analysis.*" Vol. 15, No 2, 610-625.
- [33] Nemirovskii, A. S., Polyak, B. T. and Tsybakov, A. B. (1985). "Rate of convergence of nonparametric estimates of maximum-likelihood type". *Problemy Peredachi Informatsii*, 21, 258-272.
- [34] Ramsay, J.O. (1988). "Monotone regression splines in action (with discussion)", *Statistical Science* 3(4):425-461.
- [35] Reinsch, C.R. (1967). "Smoothing by spline functions", *Numerische Mathematik* 10:177-183.
- [36] Reinsch, C.R. (1971). "Smoothing by spline functions II", *Numerische Mathematik* 16:451-454.
- [37] Schmidt, J.W. and Scholz, I.(1991). "A dual algorithm for convex-concave data smoothing by cubic  $C^2$ -splines", *Numerische Mathematik* 57:333-350.
- [38] Schwetlick, H. and Kunert, V. (1993). "Spline smoothing under constraints on derivatives", *Bit* 33:512-528.
- [39] Stone, C. J.(1982). "Optimal rates of convergence for nonparametric regression". *Ann. Statist.* 10, 1040-1053.
- [40] Tantiyaswasdikul, C. and Woodroffe, M.B. (1994). "Isotonic smoothing splines under sequential designs", *Journal of Statistical Planning and Inference* 38:75-88.
- [41] Utreras, F.I. (1985). "Smoothing noisy data under monotonicity constraints: Existence, characterization and convergence rates", *Numerische Mathematik* 47:611-625.

- [42] Van De Geer, S.(1990). "Estimating a regression function". Ann. Satist, 18, 907-924.
- [43] Wright, F. T.(1982). "Monotone regression estimates for grouped observations". Ann. Statist. 10, 278-286.
- [44] Wright, F. T.(1981). "The Asymptotic Behavior of monotone regression Estimates", Ann. Statist. Vol 9, No.2, 443-448.
- [45] Stinchcombe, M., and White, H.(1990). "Approximating and learning unknown mappings using multilayer feedforward networks with bounded weights". Preprint. San Diego: Departement of Economics, University of California.