

UNIVERSITÉ DE NEUCHÂTEL
Faculté de droit et des sciences économiques

**Estimation robuste en
contamination asymétrique
et
application aux processus
autorégressifs**

THÈSE

**présentée à la Faculté de droit et des sciences économiques
de l'Université de Neuchâtel
pour obtenir le grade de docteur ès sciences économiques**

**par
JOSEPH BARTHOULOT**



PETER LANG

**Berne · Francfort/M. · Nancy · New York
1983**

Monsieur Joseph BARTOULOT est autorisé à imprimer sa thèse de doctorat ès sciences économiques intitulée «Estimation robuste en contamination asymétrique et application aux processus autorégressifs». Il assume seul la responsabilité des opinions énoncées.

Neuchâtel, 20 juin 1983

**Le doyen
de la Faculté de droit et
des sciences économiques**

Jean-Pierre Gern

**© Editions Peter Lang SA, Berne 1983
Successesseur des Editions
Herbert Lang & Cie SA, Berne**

**Tous droits réservés. Réimpression ou reproduction interdite
par n'importe quel procédé, notamment par microfilm, xérographie,
microfiche, microcard, offset, etc.**

Impression: Weihert-Druck GmbH, Darmstadt (RFA)

Estimation robuste en contamination asymétrique et application aux processus autorégressifs

RESUME

Le but de cette recherche est l'étude du comportement des M- estimateurs d'un paramètre de translation en situation de contamination asymétrique avec application aux processus autorégressifs d'ordre 1.

Dans la première partie de ce travail, nous comparons différentes familles d'estimateurs caractérisées par des fonctions ψ symétriques et asymétriques dans diverses situations de contaminations. Comme toutes les familles d'estimateurs dépendent d'un paramètre k , nous avons comparé leurs performances sur la base de l'erreur quadratique moyenne asymptotique intégrée sur k (EQMI). Nous avons défini un deuxième critère pour dissocier les familles d'estimateurs dont les valeurs de l'EQMI sont proches; il s'agit de la stabilité de l'erreur quadratique moyenne asymptotique en fonction de k . On montre qu'aucune famille d'estimateurs n'est appropriée à toutes les formes de contamination. En particulier, les familles d'estimateurs dont la fonction ψ est symétrique conviennent mal à des contaminations asymétriques. De même les familles d'estimateurs caractérisées par une fonction ψ asymétrique sont inadaptées aux contaminations symétriques. De plus, la médiane est toujours

inappropriée, quel que soit le type de contamination. Nous proposons finalement une stratégie pour choisir une famille d'estimateurs appropriée à divers cas de contamination vraisemblables, ainsi qu'un estimateur robuste de σ afin de pouvoir calculer la valeur de k dans l'échelle des données disponibles.

La deuxième partie est consacrée à l'étude des modèles autorégressifs du premier ordre (AR(1)) lorsque les erreurs aléatoires ont une distribution gaussienne contaminée de manière asymétrique. Nous discutons les propriétés asymptotiques des M-estimateurs dans les modèles AR(1). Nous utilisons les estimateurs proposés en conclusion de la première partie pour effectuer des simulations et vérifions la convergence rapide vers les résultats asymptotiques.

Cette étude s'achève avec le traitement approfondi d'un exemple de série temporelle tiré du domaine économique.

Je tiens à remercier le professeur Michel Lejeune, directeur de thèse, qui, par ses conseils avisés et ses encouragements, m'a aidé à mener à bien cette étude. Ma reconnaissance s'adresse également à Madame H. Badan qui, par son travail minutieux de dactylographie et sa patience émérite, a réussi à donner forme à cette recherche.

TABLE DES MATIERES

RESUME	p. 5
<u>PREMIERE PARTIE</u> : ESTIMATION ROBUSTE EN CONTAMINATION ASYMETRIQUE	13
I. RAPPELS DU CADRE DE L'ESTIMATION ROBUSTE	13
I.1. Aperçus historiques et développements récents en théorie de la robustesse	13
I.1.1. Robustesse et résistance	13
I.1.2. M- estimateurs et robustesse	15
I.1.3. Variance asymptotique d'un M- estimateur	19
I.1.4. Quelques exemples de M- esti- moteurs	19
I.1.5. La courbe d'influence d'un estimateur	24
I.1.6. La courbe d'influence d'un M- estimateur	27
I.2. Problèmes d'estimation et M- estima- teurs	30
I.2.1. La non invariance d'échelle	30
I.2.2. Biais asymptotique	33
I.2.3. Procédures d'estimation	34
II. ETUDE COMPARATIVE DE M- ESTIMATEURS DE LA MOYENNE μ D'UNE LOI DE GAUSS CONTAMINEE	37
II.1. Buts de l'étude comparative	37
II.2. Estimateurs choisis	39

II.2.1. M- estimateurs dant la fonction ψ est impaire	p. 39
II.2.2. M- estimateurs dant la fonction ψ est asymétrique	40
II.3. Les modèles de contamination utilisés	45
II.3.1. Modèle de contamination exponentielle	45
II.3.2. Modèle de contamination gaussienne	50
II.4. Critères de comparaison retenus	51
II.4.1. L'écart quadratique moyen (EQM)	51
II.4.2. L'EQM intégré et la stabilité	54
II.4.3. Comparaison des familles d'estimateurs	61
II.4.4. Conclusions	105
II.5. Choix d'un estimateur robuste de dispersion	111
<u>DEUXIEME PARTIE</u> : APPLICATION AUX MODELES AUTOREGRESSIFS	117
I. ETUDE ASYMPTOTIQUE	119
I.1. Approche classique des modèles autorégressifs	119
I.1.1. Processus stochastiques	119
I.1.2. Processus stochastiques discrets stationnaires	120
I.1.3. Processus stochastiques discrets linéaires	120
I.1.4. Processus autoregressifs	122

I.1.5. Processus AR(1)	p. 123
I.1.6. Génération d'un processus AR(1) stationnaire	124
I.1.7. Estimation classique des paramètres du modèle AR(1)	125
I.1.8. Variance asymptotique des estimateurs des moindres carrés	129
I.1.9. Efficacité des estimateurs des moindres carrés	131
1.2. Types de modèles AR(1) contaminés	135
1.3. Etude théorique en situation de contamination asymétrique	138
II. SIMULATIONS POUR DES ECHANTILLONS DE TAILLE FINIE	143
III. TRAITEMENT D'UN EXEMPLE	161
ANNEXES :	
I. Calcul de la variance asymptotique des estimateurs de la moyenne d'une loi normale contaminée	179
II. Calcul du EQM asymptotique des estimateurs de l'écart-type	191
REFERENCES BIBLIOGRAPHIQUES	197

Première partie

ESTIMATION ROBUSTE EN CONTAMINATION ASYMETRIQUE

Chapitre I. RAPPELS DU CADRE DE L'ESTIMATION ROBUSTE

I.1. Aperçus historiques et développements récents en théorie de la robustesse

C'est au début des années 60 que la notion de robustesse se répandit et que les études systématiques à son propos s'intensifièrent. Peu à peu des outils d'analyse furent développés.

Les premières études systématiques entreprises en robustesse le furent par Tukey (1960) qui étudia quelques estimateurs classiques et leur substitua des alternatives simples et résistantes à des observations aberrantes (voir également Anscombe (1960), Tukey et Anscombe (1963)). Nous retraçons ici, rapidement, les définitions et résultats importants dans ce domaine et qui nous serviront par la suite.

I.1.1. Robustesse et résistance

Il n'existe pas de définition précise et unique de la robustesse car les statisticiens ont plutôt commencé par énoncer des caractéristiques que

devraient passer les procédures statistiques afin qu'on puisse les considérer comme robustes. Ainsi, bien que l'idée générale de robustesse soit la même chez tous les statisticiens, les définitions données par chacun d'eux diffèrent.

Pour Tukey (1960), la robustesse est liée à la notion d'efficacité. Un estimateur sera considéré comme robuste si son efficacité relative reste élevée pour une variété de distributions données représentant des formes de contamination diverses.

Pour Huber (1964), une mesure adéquate de la robustesse d'un estimateur est le supremum de sa variance, lorsque la distribution $F(x)$ appartient à un ensemble de lois représentant un modèle d'indétermination ou de contamination, en particulier un modèle de la forme $F(x) = (1-\epsilon)F_0(x) + \epsilon H(x)$, où ϵ et F_0 sont fixes et où H est variable et fait figure de contamination pour F_0 . Toutefois, pour des raisons de simplification mathématique de l'analyse de la robustesse, il convient mieux de prendre en compte le supremum de la variance asymptotique.

Hampel (1971) a quant à lui défini une notion de continuité. Si $X_1, X_2, \dots, X_i, \dots$ est une suite de variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.) de loi $F(x)$ et si T_n est un estimateur dont la distribution sous F est notée $\mathcal{L}_F(T_n)$, une suite $\{T_n\}$ d'estimateurs est dite robuste en F ssi :

$$\forall \epsilon > 0, \exists \delta > 0 \text{ tel que } \forall G, \forall n, \\ d(F, G) < \delta \Rightarrow d(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \epsilon$$

où d est une métrique choisie.

La robustesse définie par Huber est parfois appelée robustesse efficace car on s'intéresse à la variance d'estimateurs, alors que la robustesse définie par Hampel est parfois nommée robustesse qualitative en raison de sa généralité.

La notion de résistance est plus simple conceptuellement que celle de robustesse car elle ne se réfère pas à des distributions de probabilité et à des modèles de contamination. On dit qu'un estimateur est résistant si la valeur de l'estimation est peu sensible à un changement arbitraire (généralement petit) des valeurs des observations.

Dans la pratique de l'analyse exploratoire des données, on peut considérer que les notions de résistance et robustesse qualitative sont synonymes.

1.1.2. M- estimateurs et robustesse

Huber (1964, 1967, 1973, 1977, 1981) étudia d'abord les propriétés asymptotiques des estimateurs définis par un problème de minimum qu'il appela M- estimateurs.

Définition : M- estimateur

Soient X_1, X_2, \dots, X_n , n variables aléatoires indépendantes et identiquement distribuées de loi $F(x; \theta)$ et $T_n = T_n(X_1, \dots, X_n)$ un estimateur de θ .

Si T_n vérifie

$$\sum \rho(X_i; T_n) = \min ,$$

où ρ est une fonction donnée (généralement connexe chez Huber), alors T_n est appelé M- estimateur de θ .

En notant $\psi(x, \theta) = \frac{\partial}{\partial \theta} \rho(x; \theta)$, T_n est alors défini par $\sum \psi(X_i; T_n) = 0$.

On notera qu'en prenant $\rho(x; \theta) = -\ln f(x; \theta)$, ou $f(x; \theta) = \frac{\partial}{\partial x} F(x; \theta)$, on obtient l'estimateur du maximum de vraisemblance (MLE).

Par la suite, nous nous intéresserons uniquement au cas où θ représente un paramètre de translation d'une loi de probabilité. La fonction $F(x; \theta)$ peut alors être écrite sous la forme $F(x - \theta)$, fonction indépendante de θ .

Dans ce cas, il est normal de choisir dans la classe des M- estimateurs des estimateurs qui soient invariants par translation. Les M- estimateurs de paramètre de translation étudiés par la

suite seront donc de manière naturelle des estimateurs définis implicitement par l'équation

$$\sum \psi (X_i - T_n) = 0 \quad .$$

En posant $w_i = \frac{\psi (X_i - T_n)}{X_i - T_n}$, on obtient

$$T_n = \frac{\sum w_i X_i}{\sum w_i}$$

Par cette transformation, on remarque que T_n représente en fait une moyenne pondérée dont les poids dépendent de la distance entre l'estimateur et les variables.

A ce point, on remarque le lien qui existe entre M- estimateurs et robustesse. Supposons que la fonction $\rho(x-\theta)$ soit égale à $(x-\theta)^2$. Alors l'estimateur T_n sera défini par $\sum (X_i - T_n)^2 = \min$, soit $-2 \sum (X_i - T_n) = 0$, c'est-à-dire $T_n = \frac{1}{n} \cdot \sum X_i = \bar{X}$. Dans ce cas, $w_i = \text{constante}$, pour tout i , ce qui signifie que le poids correspondant à une variable ne dépend pas de la distance entre cette variable et T_n .

Supposons alors que nous disposons de n observations x_1, x_2, \dots, x_n parmi lesquelles certaines sont exagérément grandes ou exagérément petites (cela arrive pratiquement lorsqu'il y a mauvais fonctionnement dans une expérience, erreur de codification, etc). Ces observations seront très éloignées de la moyenne \bar{x} et elles l'influenceront donc

considérablement puisque la contribution de chaque observation x_i à \bar{x} est $\frac{x_i}{n}$, donc proportionnelle à x_i . Les différentes valeurs de \bar{x} calculées pour différents échantillons observés de taille n seront donc plus dispersées qu'elles ne le seraient sans ces observations. Ainsi l'estimateur \bar{X} perdra de la précision, qualité essentielle d'un estimateur, c'est-à-dire qu'il sera peu robuste.

Afin de diminuer la dispersion de l'estimation, l'idée est donc d'affecter des poids aux observations qui soient d'autant plus petits que l'observation est éloignée de la quantité estimée.

Affectons par exemple des poids w_i aux observations x_i qui soient inversement proportionnels à la distance en valeur absolue entre x_i et l'estimation finalement obtenue, disons \tilde{x} . Ainsi

$$w_i = \frac{1}{|x_i - \tilde{x}|}$$

$$\psi(x_i - \tilde{x}) = \frac{x_i - \tilde{x}}{|x_i - \tilde{x}|} = \begin{cases} 1 & \text{si } x_i > \tilde{x} \\ -1 & \text{si } x_i < \tilde{x} \end{cases}$$

\tilde{x} est donc la quantité caractérisée par le fait qu'il y a autant d'observations inférieures à elle que supérieures à elle : \tilde{x} est la médiane des observations. L'influence de chaque observation est indépendante de sa valeur et vaut $\frac{1}{n}$.

De manière générale, on cherche donc des

fonctions de poids $w=f(x-\theta)$ décroissantes, c'est-à-dire des fonctions $\psi(x-\theta)$ qui ne croissent pas trop rapidement, au mieux, qui soient bornées.

I.1.3. Variance asymptotique d'un M- estimateur

Huber a démontré le résultat suivant :

en écrivant $\lambda(u) = \int \psi(x;u) dF(x) = E_F[\psi(X;u)]$,

sous les hypothèses :

1) T_n converge en probabilité vers une valeur c :

$$\lambda(c) = 0$$

2) $\lambda(u)$ est différentiable au point $u=c$

3) $\int \psi^2(x;u) dF(x)$ est finie et continue au point $u=c$

alors

$$\sqrt{n} (T_n - c) \xrightarrow{\text{asympt.}} \mathcal{N}(0, \sigma^2)$$

avec

$$\sigma^2 = \frac{E_F(\psi^2(X;c))}{\left[\left(\frac{\partial}{\partial u} \lambda(u) \right)_{u=c} \right]^2}$$

Il s'agit là d'un résultat fondamental qu'il a généralisé par la suite (1967, 1981) et appliqué à la régression (1973).

I.1.4. Quelques exemples de M- estimateurs

La question qui vient naturellement à l'esprit est celle du choix d'une fonction ψ . Ce choix

dépend du but que l'on se fixe.

Pour Huber, on l'a vu plus haut, la notion de robustesse est liée à celle d'efficacité. Pour lui un estimateur robuste est un estimateur qui résiste bien à une légère modification de la spécification de la loi des variables aléatoires tout en conservant une efficacité élevée. Dans le cas de l'estimation d'un paramètre de translation θ , Huber a cherché un estimateur de variance asymptotique minimale, lorsque F est une distribution variable appartenant à un ensemble \mathcal{F} , avec $f = \frac{dF}{dx}$ symétrique. Si $V(\psi, F)$ est la variance asymptotique d'un M-estimateur défini par la fonction ψ pour la distribution F , il a alors montré (1964, p. 86 et s.) qu'il existe un point-selle (ψ_0, F_0) tel que $\sup_{F \in \mathcal{F}} V(\psi_0, F) = V(\psi_0, F_0) = \inf_{\psi \in \Psi} V(\psi, F_0)$. F_0 est la distribution pour laquelle l'information de Fisher pour θ est la plus petite et $\psi_0 = -\frac{f'_0}{f_0}$ où $f_0 = \frac{dF_0}{dx}$ est la fonction définissant l'estimateur du maximum de vraisemblance de ψ pour F_0 . En particulier, si $\mathcal{F} = \{F = (1-\epsilon)\phi_0 + \epsilon H, H \text{ symétrique}\}$ où ϕ_0 est la fonction de répartition de Gauss d'une variable aléatoire centrée et réduite, f_0 vaut

$$f_0(u) = \begin{cases} (1-\epsilon)\phi_0(-k)e^{k(u+k)} & \text{si } u \leq -k \\ (1-\epsilon)\phi_0(u) & \text{si } |u| < k \\ (1-\epsilon)\phi_0(k)e^{-k(u-k)} & \text{si } u \geq k \end{cases}$$

où $u = \frac{x-\mu}{\sigma}$ $k = k(\epsilon)$ est une fonction de ϵ , et ϕ_0

est la densité de la loi de Gauss. On en déduit que la fonction ψ_0 vaut :

$$\psi_0(u) = \begin{cases} -k & \text{si } u \leq -k \\ u & \text{si } |u| < k \\ k & \text{si } u \geq k \end{cases}$$

Les fonctions ψ_0 et f_0 sont reproduites sur la figure I.1.4.a.

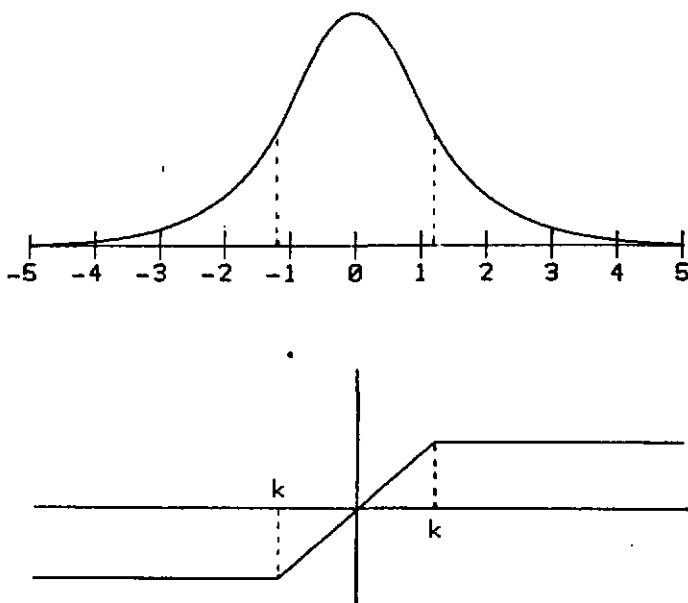


Figure I.1.4.a

Dans Andrews et autres (1972), Hampel a proposé un estimateur capable de résister à une contamination extrême représentée par des observations aberrantes d'amplitude très grande. Dans ce but la fonction ψ ne doit pas seulement être bornée mais

doit également être non monotone et se rapprocher de 0 à mesure que la variable s'éloigne du paramètre. La fonction ψ proposée est la suivante :

$$\psi(u) = \begin{cases} u & \text{si } 0 \leq |u| < a \\ \text{sign}(u) a & \text{si } a \leq |u| < b \\ \frac{a(\text{sign}(u)c-u)}{c-b} & \text{si } b \leq |u| < c \\ 0 & \text{si } c \leq |u| \end{cases}$$

L'estimateur induit par une telle fonction contient donc une règle de rejet : les observations qui sont à une distance de $\hat{\theta}$ supérieure à c seront rejetées et ne contribueront donc pas à l'estimation de θ . Cette fonction est reproduite sur la figure I.1.4.b.

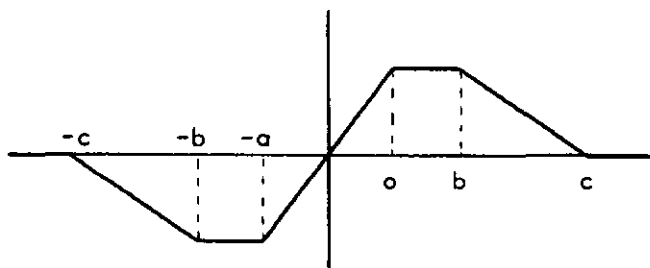


Figure I.1.4.b

Beaucoup d'autres fonctions ont été créées, par exemple le "biweight" de Tukey :

$$\psi(u) = \begin{cases} u[1-(u/a)^2]^2 & \text{si } |u| \leq a \\ 0 & \text{si } |u| > a \end{cases}$$

et le sinus de Andrews :

$$\psi(u) = \begin{cases} \sin(u/a) & \text{si } |u| \leq a \cdot \pi \\ 0 & \text{si } |u| > a \cdot \pi \end{cases}$$

Evidemment, tous les paramètres de ces fonctions dépendent de la proportion de contamination ϵ qui doit être considérée comme fixe mais inconnue en pratique. Pour choisir correctement les valeurs de ces paramètres, il est nécessaire d'étudier différentes situations de contamination qui soient vraisemblables, avec différentes valeurs de ces paramètres et ne retenir que celles qui sont les plus satisfaisantes. Une étude comparative extensive a été faite par Andrews et autres (1972). Bien qu'une certaine part de subjectivité subsiste, les valeurs suivantes paraissent raisonnables pour les paramètres des fonctions ψ données plus haut pour des situations courantes de contamination de la loi de Gauss :

Huber : $1.2 \leq k \leq 1.5$
 Hampel : $1.5 \leq a \leq 2.0$; $3.5 \leq b \leq 4.0$; $8.0 \leq c \leq 8.5$
 Tukey : $5.0 \leq a \leq 6.0$
 Andrews : $2.0 \leq a \leq 2.5$

Les estimateurs ont ainsi une efficacité asymptotique proche de 0.95 lorsque la loi de probabilité est une loi gaussienne de variance unitaire.

1.1.5. La courbe d'influence d'un estimateur

Afin de pouvoir choisir une fonction ψ qui soit pertinente pour résoudre un problème d'estimation, il est utile de pouvoir mesurer l'influence d'une variable aléatoire sur un estimateur. Pour cela, nous introduisons la notion de courbe d'influence d'un estimateur définie par Hampel. Essentiellement, cette courbe mesure l'influence marginale d'une observation sur l'estimation obtenue, si cette observation appartient à un échantillon de taille infinie.

Formellement la courbe d'influence (I.C.) est définie ainsi :

Soit F_n la fonction de répartition empirique fondée sur un échantillon aléatoire

X_1, X_2, \dots, X_n et $T_n(X) = T_n(X_1, X_2, \dots, X_n)$ un estimateur de θ fondé sur cet échantillon.

Puisqu'il y a relation biunivoque entre (X_1, X_2, \dots, X_n) et $F_n(x)$, on peut remplacer la statistique $T_n(x)$ par une autre application $T : \mathcal{F} \rightarrow R$, où \mathcal{F} est l'ensemble des fonctions en escalier, appelée fonctionnelle et qui conduit toujours au même résultat que celui obtenu par $T_n(X)$. On peut donc écrire l'égalité suivante :

$$T_n(X) = T_n(X_1, X_2, \dots, X_n) = T(F_n) \quad .$$

Lorsque n tend vers l'infini, F_n converge uniformément en x vers F et $T(F_n)$ converge en général vers $T(F)$. Alors la courbe d'influence de T pour F est définie par :

$$IC(x; T, F) = \lim_{t \rightarrow 0} \frac{T[(1-t)F + t\delta_x] - T(F)}{t}$$

où δ_x est la fonction de Dirac en x .

Exemples de courbes d'influence de fonctionnelles linéaires

a) moyenne : $F \xrightarrow{T_1} T_1(F) = \int x dF(x) = \mu$

$$IC(x; T_1, F) = \lim_{t \rightarrow 0} \frac{(1-t)\mu + tx - \mu}{t} = x - \mu$$

b) variance : $F \xrightarrow{T_2} T_2(F) = \int (x - \mu)^2 dF(x) = \sigma^2$

$$IC(x; T_2, F) = \lim_{t \rightarrow 0} \frac{(1-t)\sigma^2 + t(x - \mu)^2 - \sigma^2}{t} = (x - \mu)^2 - \sigma^2$$

On remarque que l'influence sur T_1 d'un ajout infinitésimal en un point x quelconque est proportionnelle à la distance entre x et μ .

De même, l'influence sur T_2 d'un ajout infinitésimal en un point x quelconque est proportionnelle au carré de la distance entre x et μ . Si $|x - \mu| < \sigma$, l'influence sera négative. Cela signifie

que l'ajout tend à diminuer la variance. Par contre, si $|x-\mu| > \sigma$, l'ajout tend à augmenter la variance sans limites. $IC(x;T_1,F)$ et $IC(x;T_2,F)$ ont été reproduites sur la figure I.1.5.a.

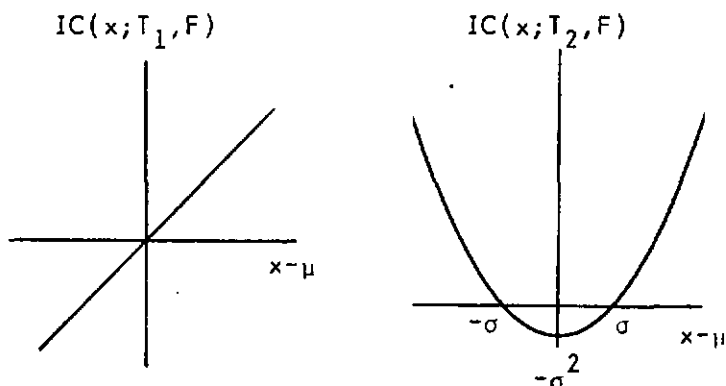


Figure I.1.5.a

La représentation de la nature de courbe d'influence, pour un échantillon de taille finie peut se faire de la manière suivante, pour la moyenne et la variance.

Supposons que l'on dispose de n observations x_1, x_2, \dots, x_n , réalisation d'un échantillon aléatoire X_1, X_2, \dots, X_n . De plus, on ajoute une observation en un point quelconque x . L'influence de cette observation sur la moyenne empirique \bar{x}_n sera égale à :

$$\bar{x}_{n+1} - \bar{x}_n = \frac{n \bar{x}_n + x}{n+1} - \bar{x}_n = \frac{x - \bar{x}_n}{n+1}$$

c'est-à-dire que l'influence sera proportionnelle à la distance entre x et \bar{x}_n et inversement proportion-

nelle à la taille de l'échantillon.

L'influence de cette même observation sur $s_n^2 = \frac{\sum (x_i - \bar{x})^2}{n}$, estimation de la variance, vaudra

$$s_{n+1}^2 - s_n^2 = \frac{ns_n^2 + (x - \bar{x}_n)^2}{n+1} - s_n^2 = \frac{(x - \bar{x}_n)^2 - s_n^2}{n+1}$$

Si $|x - \bar{x}_n| < s_n$, la variance diminuera, mais au plus de $\frac{s_n^2}{n+1}$. Par contre si $|x - \bar{x}_n| > s_n$ alors la variance augmentera très vite et cela sans limite.

1.1.6. Courbe d'influence d'un M- estimateur

On trouve de manière directe (Huber (1977, 1981)) la courbe d'influence d'un M- estimateur, dans le cas général :

en posant : $F_t = (1-t)F_0 + t\delta_x$, la courbe d'influence de la fonctionnelle T en F_0 peut s'écrire :

$$IC(x; T, F_0) = \left. \frac{d}{dt} T(F_t) \right|_{t=0}$$

Pour un M- estimateur T_n induit par la fonctionnelle T définie par

$$\int \psi(x; T(F)) dF(x) = 0, \quad \text{I.1.6.o}$$

on obtient la courbe d'influence en remplaçant F par F_t et en dérivant cette expression par rapport à t , soit :

$$\begin{aligned} \frac{d}{dt} \int \psi(x; T(F_t)) dF_t(x) \Big|_{t=0} &= 0 \\ \int \frac{\partial}{\partial T(F_t)} \psi(x; T(F_t)) \cdot \frac{d}{dt} T(F_t) \cdot dF_t(x) \Big|_{t=0} &+ \\ + \int \psi(x; T(F_t)) \cdot d(\delta_x - F_0)(x) \Big|_{t=0} &= 0 \\ \frac{d}{dt} T(F_t) \Big|_{t=0} \cdot \int \frac{\partial}{\partial T(F_0)} \psi(x; T(F_0)) dF_0(x) &+ \\ + \psi(x; T(F_0)) - 0 &= 0 \end{aligned}$$

soit,

$$IC(x; T, F_0) = \frac{\psi(x; T(F_0))}{-\int \frac{\partial}{\partial T(F_0)} \psi(x; T(F_0)) dF_0(x)}$$

c'est-à-dire que la courbe d'influence est proportionnelle à la fonction ψ .

Propriétés diverses

- 1) $E[IC(X; T, F)] = 0$, en raison de la relation I.1.6.a, donc
- 2) $V[IC(X; T, F)] = \int (I.C(x; T, F))^2 dF(x)$.

On remarque que la variance de la courbe d'influence qui est égale à

$$\frac{\int \psi^2(x; T, (F)) dF(x)}{\left(\int \frac{\partial}{\partial T(F)} \psi(x; T(F)) dF(x)\right)^2}$$

est la même que la variance asymptotique de $\sqrt{n} T(F_n)$ donnée sous I.1.3.; mis à part l'inversion des opérations d'intégration et de dérivation ou dénominateur.

3) $T(F_n)$ sera efficace asymptotiquement ssi

$$\frac{E[(IC(X; T, F))^2]}{\frac{1}{I(F)}} = 1$$

où $I(F) = \int \frac{\partial}{\partial \theta} \ln f(x; \theta) dF(x; \theta)$ est l'information de Fisher pour θ contenue dans la loi $f(x; \theta)$ c'est-à-dire si

$$IC(x; T, F) = \frac{1}{I(F)} \cdot \frac{\partial}{\partial \theta} \ln f(x; \theta)$$

soit, pour un M-estimateur d'un paramètre de translation, si

$$\psi(x - T(F)) = - \frac{C}{I(F)} \cdot \frac{\partial}{\partial \theta} \ln f(x - \theta)$$

où $C = \int \frac{\partial}{\partial T(F)} \psi(x - T(F)) dF(x)$

c'est-à-dire si l'estimateur est proportionnel au MLE.

1.2. Problèmes d'estimation et M- estimateurs

1.2.1. La non invariance d'échelle

Les M- estimateurs ne sont en général pas invariants d'échelle. En effet, soit T_n estimateur du paramètre de translation θ et solution de l'équation $\Sigma \psi(X_i - T_n) = 0$ et $\lambda \in \mathbb{R}$. $\lambda \cdot T_n$ ne sera solution de l'équation $\Sigma \psi(\lambda X_i - \lambda T_n) = 0$ quelles que soient les valeurs que prendront les variables aléatoires X_i que si ψ est une fonction homogène. Or en général, les fonctions ψ ne le sont pas et on le vérifie aisément par exemple dans le cas de la fonction ψ de Huber :

$$\psi(X_i - T_n) = \begin{cases} -k & \text{si } X_i - T_n \leq -k \\ X_i - T_n & \text{si } |X_i - T_n| < k \\ k & \text{si } X_i - T_n \geq k \end{cases}$$

Autrement dit si on multiplie chaque x_i par un nombre λ , l'estimation obtenue ne sera pas $\lambda \hat{\theta}$. Il faut donc transformer notre estimateur afin de le rendre invariant d'échelle (on parle également à juste titre d'équivariance plutôt que d'invariance). Une façon simple d'y arriver est de transformer les variables de manière à toujours travailler avec des grandeurs comparables. Pour cela, il suffit de réduire les variables, c'est-à-dire les diviser par l'écart-type σ_x . Par exemple l'estimateur de Huber sera rendu invariant d'échelle s'il est solution de l'équation $\Sigma \psi\left(\frac{X_i - T_n}{\sigma_x}\right) = 0$ avec ψ défini par :

$$\psi\left(\frac{X_i - T_n}{\sigma_x}\right) = \begin{cases} -k & \text{si } \frac{X_i - T_n}{\sigma_x} \leq -k \\ \frac{X_i - T_n}{\sigma_x} & \text{si } -k < \frac{X_i - T_n}{\sigma_x} < k \\ k & \text{si } \frac{X_i - T_n}{\sigma_x} \geq k \end{cases} \quad (I.2.1.a)$$

Donc vouloir estimer un paramètre de translation à l'aide des M-estimateurs ne se réduit pas à un problème pur d'estimation d'un paramètre de translation. Il intervient le paramètre σ_x généralement inconnu, donc à estimer, et qu'on peut considérer comme un paramètre de nuisance à notre problème d'estimation.

Finalement T_n estimateur de θ , invariant par translation et d'échelle, sera solution de l'équation $\sum \psi\left(\frac{X_i - T_n}{S_n}\right) = 0$. S_n estimateur de σ_x doit également posséder des propriétés de robustesse afin de ne pas annihiler les propriétés de robustesse de T_n conférées par la fonction ψ . En effet, en estimant σ_x par $S_n = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$, S_n prendra des valeurs très grandes lorsque des variables X_i prendront des valeurs éloignées de \bar{X}_n . Les variables $\frac{X_i - T_n}{S_n}$ seront donc exagérément réduites. Alors, en utilisant par exemple la fonction ψ de Huber donnée en (I.2.1.), on remarque que le nombre de variables telles que $\left|\frac{X_i - T_n}{S_n}\right| < k$

sera supérieur à ce que l'on aurait souhaité et ainsi le but qui était de diminuer le poids des variables éloignées de T_n ne sera pas atteint.

Il existe différentes méthodes d'estimation de σ qu'on peut grouper en deux catégories :

- a) Les méthodes d'estimation simultanées de θ et σ , à savoir trouver les solutions T_n et S_n des équations

$$\begin{aligned} \sum \psi\left(\frac{X_i - T_n}{S_n}\right) &= 0 \\ \sum X\left(\frac{X_i - T_n}{S_n}\right) &= 0 \end{aligned}$$

Huber (1964, 1981) a proposé de telles méthodes.

- b) Les méthodes d'estimation indépendantes, à savoir trouver T_n solution de l'équation $\sum \psi\left(\frac{X_i - T_n}{S_n}\right) = 0$ et estimer σ par un estimateur S_n indépendant de T_n choisi arbitrairement, mais robuste.

Nous avons adopté la deuxième solution car cette étude est d'abord centrée sur l'estimation robuste des paramètres de translation. Ensuite, les résolutions numériques sont plus simples lorsqu'on estime indépendamment les paramètres θ et σ . Enfin, comme nous allons procéder à des comparaisons des performances d'estimateurs de paramètre de translation, la méthode d'estimation de σ aura peu d'importance, pourvu qu'on applique la même méthode à chaque estimateur de θ .

1.2.2. Biais asymptotique

Le but est d'estimer le paramètre de translation d'une loi donnée qui n'est pas nécessairement un centre de symétrie (par exemple le paramètre θ de la loi uniforme $U_{[\theta, \theta+1]}$ est un paramètre de translation) et ce paramètre est estimé par un M-estimateur qui représente une moyenne pondérée.

Mais il faut garder à l'esprit que la loi pour laquelle on désire estimer le paramètre de translation n'est pas connue de manière exacte, mais appartient à un ensemble $\mathcal{F} = \{F = (1-\epsilon)G + \epsilon H\}$. Afin que l'estimateur T_n converge vers θ , il est nécessaire que $\int \psi(x-\theta) dF(x; \theta) = 0$ pour tout $F \in \mathcal{F}$. Une condition suffisante pour y arriver est de restreindre \mathcal{F} à un ensemble de fonctions telles que $\forall F \in \mathcal{F}$, $\frac{dF}{dx}$ est symétrique et restreindre l'ensemble Ψ des fonctions ψ à un ensemble de fonctions impaires, le paramètre à estimer étant le centre de symétrie.

Restreindre \mathcal{F} à un ensemble de lois symétriques peut être obtenu en imposant que la loi $\frac{dG}{dx}$ est symétrique et les lois $\frac{dH}{dx}$ sont également symétriques. Cela signifie que la contamination se répartit symétriquement relativement au centre de symétrie. Un tel modèle de contamination est simple théoriquement et s'interprète facilement : un nombre trop élevé d'observations dans un échantillon se situent trop loin (valeurs aberrantes) et/ou trop près (contamination cachée) du centre de symétrie.

C'est généralement cette hypothèse qui a été adoptée jusqu'à présent dans les études des M- estimateurs robustes des paramètres de translation.

Il est alors naturel d'estimer le centre de symétrie comme paramètre de translation.

Ce qui est gênant en adoptant de telles hypothèses c'est qu'on transgresse l'idée fondamentale sous-jacente à l'estimation robuste qui est de pouvoir estimer des paramètres (de translation ou d'échelle) à partir d'échantillons entachés d'une contamination inconnue et donc éventuellement asymétrique. Dans la suite de cette étude, nous nous intéresserons au cas général où la symétrie des fonctions F n'est plus vérifiée. Collins (1976) et Jaeckel (1971) ont également étudié les situations de contamination asymétrique, mais par des cheminements plus théoriques que celui que nous adapterons ici.

I.2.3. Procédures d'estimation

Calculer la moyenne ou la médiane d'un échantillon requiert peu d'habileté dans la manipulation des chiffres car les opérations arithmétiques à effectuer sont simples. Par contre, estimer une valeur fondée sur un M- estimateur dont la fonction ψ n'est pas triviale peut exiger un grand nombre d'opérations à effectuer et le recours à l'ordinateur est souvent nécessaire. En effet, puisqu'on peut écrire

$T_n = \frac{w_i X_i}{w_i}$ et que w_i est une fonction de T_n , la solution d'une telle équation devra se trouver par une procédure itérative. Ce genre de problèmes est bien connu et il existe différents algorithmes pour les résoudre, par exemple

$$o) \text{ Newton : } \hat{\theta}_{j+1} = \hat{\theta}_j + \frac{s \sum_i \psi[(x_i - \hat{\theta}_j)/s]}{\sum_i \psi'[(x_i - \hat{\theta}_j)/s]}$$

$$b) \text{ moyenne pondérée : } \hat{\theta}_{j+1} = \frac{\sum_i w_{ij} x_i}{\sum_i w_{ij}}$$

$$\text{avec } w_{ij} = \frac{\psi[(x_i - \hat{\theta}_j)/s]}{(x_i - \hat{\theta}_j)/s}$$

s est une estimation robuste de σ .

Lorsque la fonction ψ n'est pas monotone (cas de celle proposée par Hampel ou Tukey), l'équation à résoudre peut admettre plusieurs solutions. Afin de se prémunir de telles éventualités, il est nécessaire de prendre certaines précautions et d'adopter des conventions (voir Collins (1976)) :

- On partira en première estimation de la médiane plutôt que de la moyenne.
- Si la suite $\hat{\theta}_j$ des estimations intermédiaires n'est pas convergente, on arrêtera les calculs et on retiendra la médiane comme estimation finale.

Chapitre II. ETUDE COMPARATIVE DE M- ESTIMATEURS DE LA MOYENNE μ D'UNE LOI DE GAUSS CONTAMINEE

II.1 Buts de l'étude comparative

Nous nous intéressons maintenant au cas où la contamination est asymétrique. Le modèle de contamination considéré sera de la forme $\mathcal{F} = \{F = (1-\epsilon)\phi + \epsilon H\}$ où ϕ est la distribution de Gauss de paramètres μ pour la moyenne et σ^2 pour la variance.

Si T_n converge vers μ pour un F_0 particulier, il ne convergera pas en général vers μ pour tout $F \in \mathcal{F}$, comme on l'a vu en I.2.b, car la condition $\int \psi(x-\mu) dF(x) = 0$, $\forall F \in \mathcal{F}$ ne sera pas vérifiée.

De plus, puisque F est variable, il n'est pas possible de trouver un estimateur ayant des propriétés optimales pour tout $F \in \mathcal{F}$.

Supposons en effet le modèle de contamination suivant : $\mathcal{F} = \{F = (1-\epsilon)\phi(x; \mu_1, \sigma_1) + \epsilon\phi(x; \mu_2, \sigma_2)\}$. Pour ϵ proche de 0 on considérera que la contamination provient du deuxième terme et donc le paramètre que l'on voudra estimer est μ_1 . Pour un estimateur T_n donné la validité de l'estimation dépendra de la proportion de contamination ϵ fixe mais inconnue, ainsi que des paramètres μ_2 et σ_2 de la loi perturbatrice.

Par exemple, si l'estimateur choisi est \bar{X}_n (la moyenne arithmétique des observations), il sera sans biais pour μ_1 si $\epsilon=0$ ou si $\mu_2=\mu_1$. Dans le premier cas \bar{X}_n sera MVUE, mais pas dans le second car la loi considérée n'est plus la même (sauf si $\sigma_2=\sigma_1$).

Donc le but n'est pas de trouver un estimateur possédant des propriétés optimales pour une loi particulière, mais bien un estimateur montrant des propriétés satisfaisantes pour un ensemble de lois considérées. Dans ce but, nous allons donc comparer le comportement asymptotique de divers estimateurs dans plusieurs situations de contamination. Puisque les estimateurs auront en général un biais, il paraît naturel de prendre comme critère de qualité l'écart quadratique moyen (EQM) asymptotique.

Pour une loi donnée, comme l'EQM est une fonction du paramètre à estimer, on ne peut pas dans le cas général trouver d'estimateur qui ait un EQM inférieur ou égal à celui de tout autre estimateur, pour toutes les valeurs que peut prendre le paramètre à estimer. Un exemple bien connu est le suivant :

Si X est une v.a. distribuée par une loi binomiale $\mathcal{B}(n, p)$, $\frac{X}{n}$ est un estimateur sans biais de p . En prenant l'estimateur $\frac{X+1}{n+2}$ on obtient un estimateur biaisé de p , sauf si $p=1/2$. Par contre $EQM(\frac{X+1}{n+2}) < EQM(\frac{X}{n})$ pour $p(1-p) > \frac{1}{8}$ et $EQM(\frac{X+1}{n+2}) > EQM(\frac{X}{n})$ pour $p(1-p) < \frac{1}{8}$ lorsque n tend vers l'infini. Autrement dit $\frac{X+1}{n+2}$ est plus précis (respect. moins précis) que $\frac{X}{n}$ lorsque p

est proche de $1/2$ (respect. éloigné de $1/2$).

Toutefois, par construction nous évitons ce problème puisque nous allons estimer un paramètre de translation par des estimateurs invariants par translation, et qu'alors si $EQM(T_n) < EQM(T'_n)$ pour une valeur particulière θ_0 appartenant à l'espace θ du paramètre, il le sera également pour toute valeur $\theta \in \theta$ pour la famille de lois considérées $F(x; \theta)$. En effet, en faisant subir à chaque X_i une translation $Y_i = X_i + m$, et en écrivant $X = (X_1, X_2, \dots, X_n)$ et $Y = (Y_1, Y_2, \dots, Y_n)$, alors $T_n(Y) = T_n(X) + m$ par construction des estimateurs invariants par translation et

$$E_{\theta_0} [(T_n(X) - \theta_0)^2] = E_{\theta'_0} [(T_n(Y) - \theta'_0)^2]$$

avec $\theta'_0 = \theta_0 + m$ et où les indices θ_0 et θ'_0 de E indiquent de quelle loi particulière appartenant à la famille $F(x; \theta)$ considérée l'échantillon provient.

II.2 Estimateurs choisis

Nous avons sélectionné deux types de M-estimateurs.

II.2.1 M-estimateurs dont la fonction ψ est impaire

La fonction ψ est définie ainsi :

$$\psi(u) = \begin{cases} \text{sign}(u) \gamma k (\alpha + \beta) & \text{si } |u| \geq k \\ \alpha u + \text{sign}(u) \beta k & \text{si } |u| < k \end{cases}$$

où $u = \frac{x-c}{\sigma_x}$, c défini implicitement

$$\text{par } \int \psi\left(\frac{x-c}{\sigma_x}\right) dF(x) = 0$$

$$\sigma_x^2 = \int (x-\mu_x)^2 dF(x) \quad \text{et}$$

$$\mu_x = \int x dF(x) \quad .$$

Ce choix d'estimateurs permet de couvrir une large gamme des estimateurs proposés dans la littérature pour le cas de la contamination symétrique.

Les estimateurs de ce type dépendent des paramètres α, β, γ et k . Lorsque k est considéré comme variable, les paramètres α, β, γ définissent une famille d'estimateurs que nous noterons par la suite $S(\alpha; \beta; \gamma)$ (S pour symétrique).

II.2.2. M- estimateurs dont la fonction ψ est asymétrique

La fonction ψ est définie ainsi :

$$\psi(u) = \begin{cases} u & \text{si } u \leq 0 \\ \alpha u + \beta k & \text{si } 0 < u \leq k \\ \gamma k(\alpha + \beta) & \text{si } u > k \end{cases}$$

Par la suite les familles d'estimateurs de ce type seront écrites symboliquement $A(\alpha; \beta; \gamma)$ (A pour asymétrique).

Parmi chacun des deux types d'estimateurs, nous avons choisi trois genres de familles, chaque

genre étant un type d'estimateurs pour lequel un ou deux des trois paramètres α, β, γ sont fixes, et pour chaque genre nous avons sélectionné une ou plusieurs familles d'estimateurs.

Le tableau II.2.a résume les familles d'estimateurs choisies et indique les numéros des figures sur lesquelles on a dessiné les fonctions ψ de ces familles d'estimateurs.

Remarques :

- 1) Parmi les 18 familles d'estimateurs retenues, on reconnaît certains estimateurs classiques tels la médiane ($S(0;1;1)$) indépendante de k ; la famille d'estimateurs proposée par Huber ($S(1;0;1)$) proche de la moyenne α -tranquée avec $\alpha=F(-k)$, la famille $S(1;0;1.5)$ proche d'une moyenne windsorisée, ainsi que la famille $S(1;0;0)$, moyenne "coupée", proposée également par Huber.
- 2) La forme des fonctions ψ asymétriques a été choisie relativement à une contamination supposée positive. Il est évident que par symétrie on obtiendrait les mêmes estimateurs pour des contaminations négatives.
- 3) On remarque que parmi l'ensemble des familles d'estimateurs présentées ici, deux familles possèdent une règle de rejet des observations : $S(1;0;0)$ et $A(1;0;0)$.

Les familles d'estimateurs dépendent du paramètre k et seront étudiées en fonction de k . Cela a l'avantage de voir comment les estimateurs d'une famille varient en fonction de k et de proposer un choix pour k .

types	classes	familles	figures
$S(\alpha; \beta; \gamma)$	$S(\alpha; 0; 1)$	$\alpha=1$	II.2.1.1.a
	$S(1; 0; \gamma)$	$\gamma=0; 0.5; 1.5$	II.2.1.1.b
	$S(\alpha; 1-\alpha; 1)$	$\alpha=0; 0.25; 0.5; 0.75;$	II.2.1.1.c
$A(\alpha; \beta; \gamma)$	$A(\alpha; 0; 1)$	$\alpha=0.5; 1.0; 1.5$	II.2.2.1.a
	$A(1; 0; \gamma)$	$\gamma=0; 0.5; 1.5$	II.2.2.1.b
	$A(\alpha; 1-\alpha; 1)$	$\alpha=0; 0.25; 0.5; 0.75;$	II.2.2.1.c

Tableau II.2.a

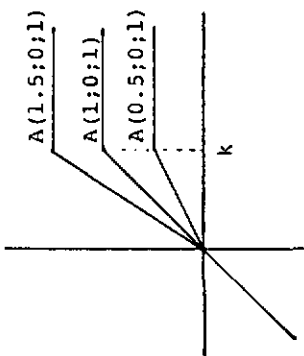


Figure II.2.2.a

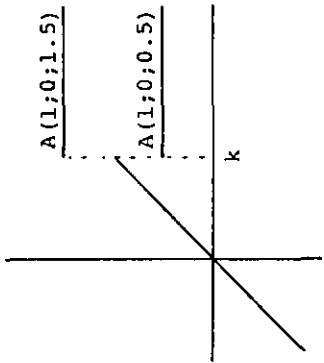


Figure II.2.2.b

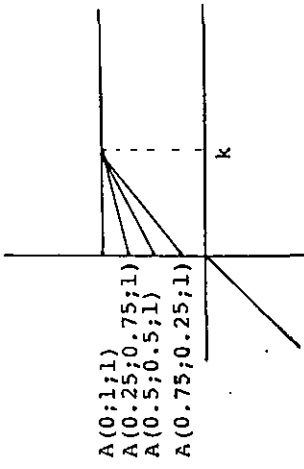


Figure II.2.2.c

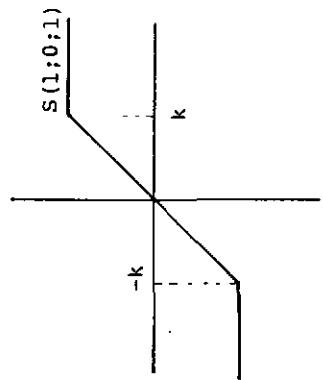


Figure II.2.1.a

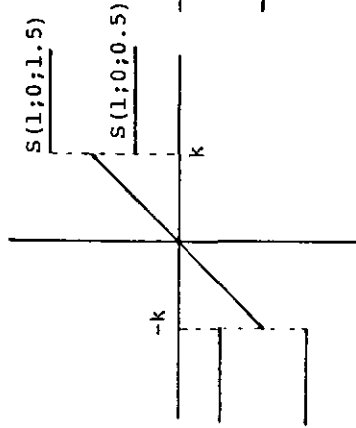


Figure II.2.1.b

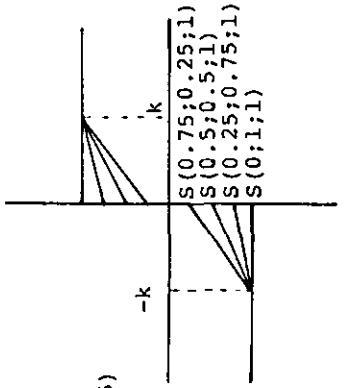


Figure II.2.1.c

11.3. Les modèles de contamination utilisés

Nous avons retenu deux types de modèles de contamination. Les résultats seront toujours décrits pour ces deux types.

11.3.1. Modèle de contamination exponentielle

La fonction de densité de la v.a. X est supposée être de la forme :

$$h(x) = (1-\varepsilon)\varphi_0(x)I_{]-\infty; +\infty[} + \varepsilon g(x)I_{[a; +\infty[}$$

où $\varphi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ représente la densité de la loi normale centrée et réduite et

$$g(x) = \frac{\mu}{\varepsilon} e^{-\lambda(x-a)}I_{[a; +\infty[} - \frac{(1-\varepsilon)}{\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}I_{[a; +\infty[}$$

Les conditions à respecter sont :

a) $g(a) = 0 \Rightarrow \mu = (1-\varepsilon)\varphi_0(a)$

b) $g(x)$ est une densité, donc

$$\int_a^{\infty} g(x)dx = 1 \Rightarrow \lambda = \frac{\mu}{1 - (1-\varepsilon)\Phi(a)}$$

où $\Phi_0(x) = \int_{-\infty}^x \varphi_0(u)du$

On a exprimé μ et λ en fonction de ε et a car pour choisir des situations de contamination, il est plus intéressant de faire varier librement

ces deux derniers paramètres qui représentent respectivement la proportion de contamination et le point à partir duquel la contamination commence.

c) $g(x)$ est une densité, donc

$$x \in [a, \infty[, g(x) \geq 0$$

On doit donc avoir :

$$\mu e^{-\lambda(x-a)} \geq \frac{1-x}{2} e^{-\frac{x^2}{2}} \quad x \geq a, \text{ soit}$$

$$-\frac{1}{2}a^2 + \lambda a - \lambda x + \frac{1}{2}x^2 \geq 0 \quad x \geq a,$$

ce que l'on peut considérer comme une fonction de k . Cette fonction s'annule pour $k = \frac{-\lambda \pm \sqrt{\lambda^2 - 1}}{-1}$.

Comme $f(a) \geq 0 \quad x \geq a$, on en déduit que

$$\{a \mid x \geq a, f(a) \geq 0\} = \{a \mid a \geq \lambda\}$$

$$\text{Puisque } \lambda = \frac{\mu}{1 - (1-\epsilon)\phi(a)}$$

$$\epsilon = 1 - \frac{1}{\frac{\phi_a(a)}{\lambda} + \phi_0(a)}$$

et comme $k \geq \lambda$

$$1 \geq \epsilon \geq 1 - \frac{1}{\frac{\phi_a(a)}{a} + \phi_0(a)}$$

On obtient donc finalement pour le choix des couples (a, ϵ) le domaine au-dessus de la courbe dessiné sur la figure 11.3.1.a.

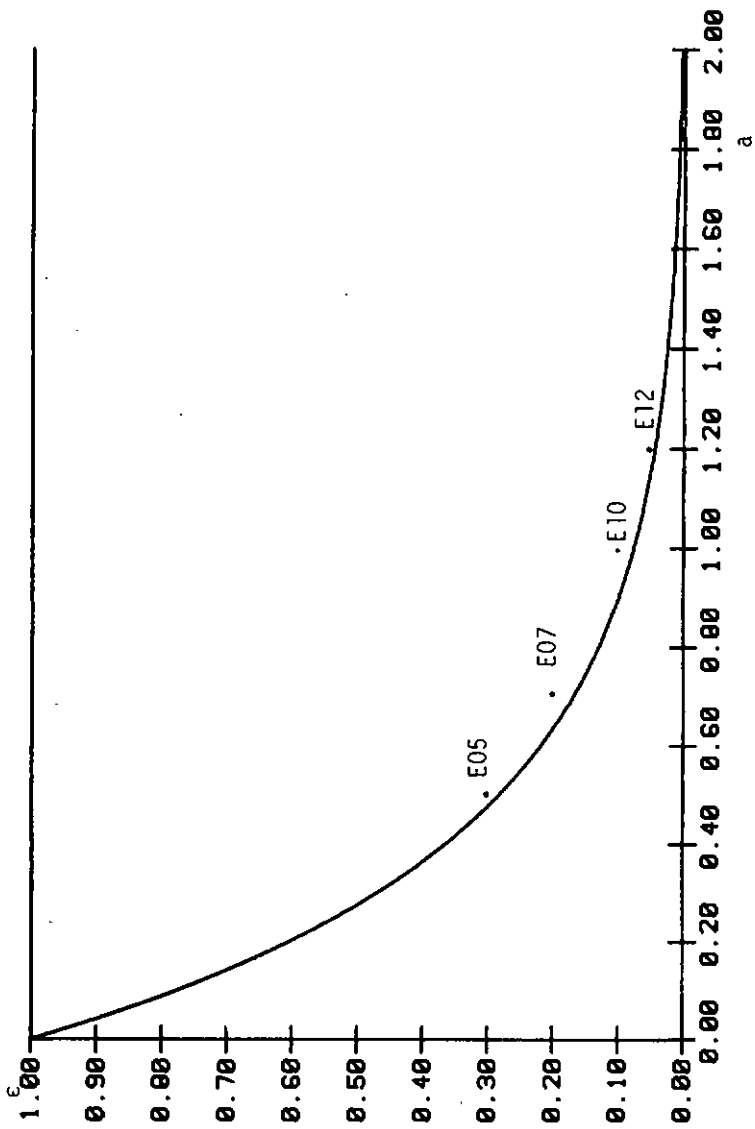


Figure II.3.1.a

Afin d'obtenir des densités ayant un aspect assez lissé, nous choisirons des couples (ϵ, a) proches de la courbe délimitant le domaine, dans sa partie inférieure.

Nous retiendrons les 4 couples suivants :

<u>epsilon</u>	<u>a</u>	<u>nom symbolique</u>	<u>interprétation</u>
0.05	1.2	E12	contamination faible
0.1	1.0	E10	contamination moyenne
0.2	0.7	E07	contamination élevée
0.3	0.5	E05	contamination très élevée

La figure II.3.1.b montre les densités correspondantes. Sur chaque dessin se trouvent $h(x) = \varphi_0(x)$, c'est-à-dire la densité non contaminée et $h(x) = (1-\epsilon)\varphi_0(x) + \epsilon g(x)$.

Ce modèle a l'avantage d'être simple dans sa conception. Il correspond à une densité non contaminée jusqu'à un point a (densité de la forme $\exp(-x^2)$) et à une densité contaminée à partir du point a (densité de la forme $\exp(-x)$) qui a pour effet d'allonger les queues de la loi; un choix judicieux des couples (ϵ, a) permet d'obtenir une densité d'aspect suffisamment lissé. Toutefois, ce modèle présente les inconvénients suivants :

- On a déjà vu qu'il n'était pas possible dans un tel modèle de choisir n'importe quel couple (ϵ, a) . Notamment, il est impossible de produire une contamination symétrique. Or, une telle contamination est

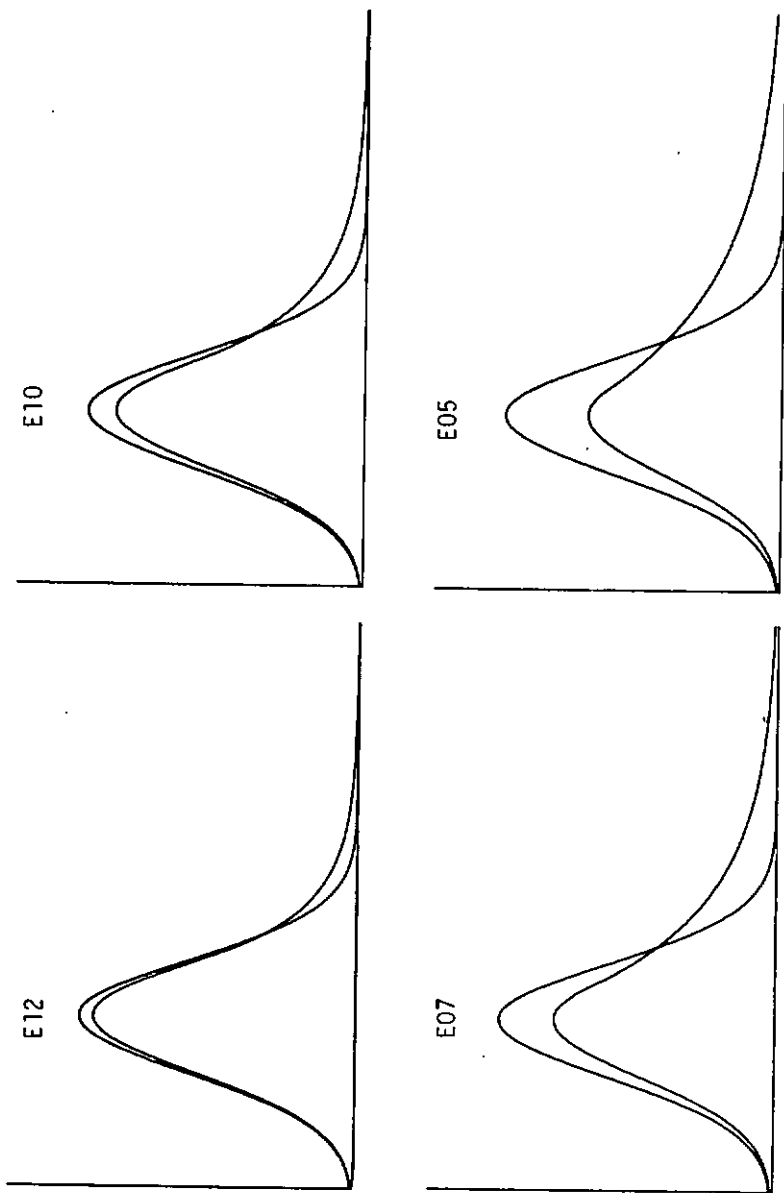


Figure II.3.1.b

intéressante, car elle permet de comparer le comportement des différents estimateurs retenus pour un ensemble plus large de situations.

- La fonction de densité n'est pas dérivable au point a , ce qui complique le calcul de la variance asymptotique des estimateurs.
- Enfin, il est impossible également d'obtenir une densité bimodale.

II.3.2. Modèle de contamination gaussienne

La fonction de densité de la v.a. X est une combinaison convexe de deux lois normales :

$$h(x) = (1-\varepsilon)\varphi_0(x) + \varepsilon\varphi_1(x)$$

où $\varphi_0(x)$ représente comme auparavant la densité de la loi normale centrée, réduite, et

$$\varphi_1(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{est la densité}$$

d'une v.a. de loi normale, de moyenne μ et de variance σ^2 .

Ce modèle permettra de générer des contaminations asymétriques unimodales, asymétriques bimodales et symétriques. Parmi l'ensemble des triplets $(\mu, \sigma, \varepsilon)$ possibles, nous avons retenu les situations suivantes :

μ	σ	ε	nom symbolique	Interprétation
0	4	0.05	G005	faible contamination symétrique
0	4	0.1	G010	contamination symétrique moyenne
0	1	0	G000	pas de contamination
2	1	0.05	G205	faible contamination asymétrique
2	1	0.1	G210	contamination asymétrique moyenne
4	1	0.05	G405	contamination asymétrique éloignée
4	1	0.1	G410	contamination asymétrique éloignée avec apparition d'un 2ème mode

Les densités représentant ces situations sont dessinées sur les figures 11.3.2.a et 11.3.2.b.

11.4. Critères de comparaison retenus

11.4.1. L'écart quadratique moyen (EQM)

Le critère retenu pour juger de la qualité d'un estimateur est, nous l'avons déjà dit, l'EQM asymptotique. Toutefois, ce critère seul est inapproprié, car dans une situation de contamination donnée, un estimateur T_n d'une certaine famille peut avoir un EQM asymptotique inférieur à celui d'un estimateur T'_n d'une autre famille pour une valeur k_0 de k et un EQM asymptotique supérieur pour une autre valeur k_1 de k . Etant donné l'infinité des valeurs de k qu'il

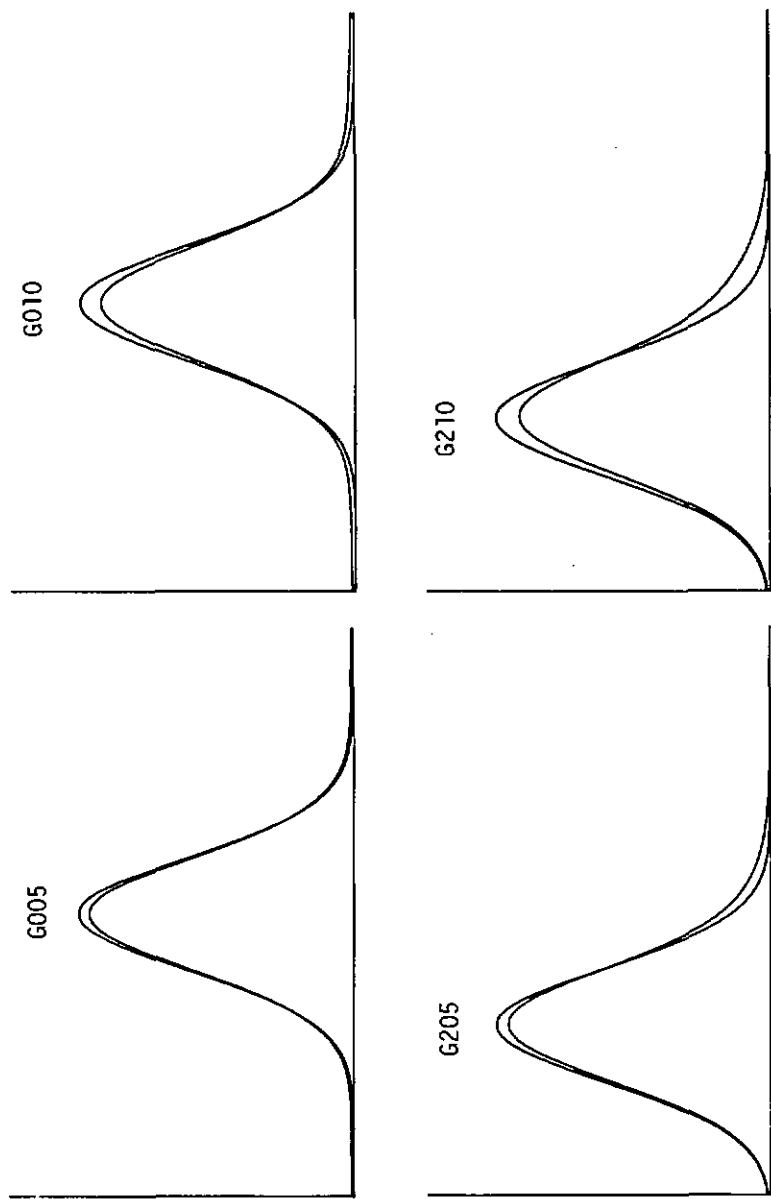


Figure II.3.2.d

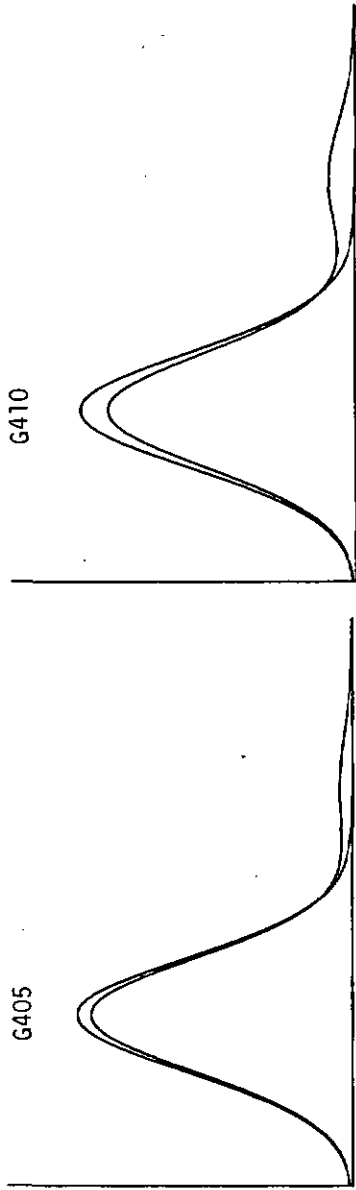


Figure II.3.2.b

est possible d'affecter à une famille d'estimateurs et, étant donné le nombre de familles d'estimateurs et de situations de contamination sélectionnées, il est impensable et stérile de vouloir comparer les estimateurs entre eux. Par contre, il est intéressant de pouvoir comparer des familles d'estimateurs. Il faut donc trouver un critère global qualifiant une famille d'estimateurs en éliminant le rôle de k . Pour cela on calculera la valeur moyenne de l'EQM asymptotique de T_n sur un intervalle $[k_1, k_2]$ de valeurs utiles de k .

II.4.2. L'EQM intégré et la stabilité

Définition :

Soit $T_n(k)$ un estimateur d'une famille donnée T_n correspondant à une valeur k du paramètre, et $EQM[T(k)]$ son écart quadratique moyen asymptotique. On définit l'EQM asymptotique intégré de la famille d'estimateurs T_n par :

$$EQMI(T) = \frac{\int_{k_1}^{k_2} EQM[T(k)] dk}{k_2 - k_1}$$

Les bornes k_1 et k_2 doivent être choisies en tenant compte des éléments suivants :

- a) Il faut qu'entre ces bornes se trouvent les valeurs souvent recommandées en pratique pour les estimateurs connus (par exemple Andrews et autres

(1971)).

- b) En pratique, σ n'est pas connu et devra être estimé afin de pouvoir en déduire la valeur de k dans l'échelle effective des données, à savoir $k\sigma$. Donc $k\sigma$ devra être estimé par ks où s est une estimation robuste de σ . Il faut donc choisir un intervalle $[k_1, k_2]$ qui puisse refléter les variations vraisemblables de ks .
- c) Pour toutes (ou presque) les situations de contamination retenues, l'estimateur qui dans chaque famille a l'EQM asymptotique le plus faible devrait avoir une valeur de k comprise entre k_1 et k_2 .
- d) Il faut que la discrimination entre familles d'estimateurs soit grande afin de pouvoir procéder à des comparaisons sensées. Comme le critère EQMI est une moyenne sur l'intervalle $[k_1, k_2]$, il faut donc que cet intervalle soit petit afin d'augmenter les différences entre moyennes.

Satisfaire simultanément ces 4 conditions nous a conduits à choisir les valeurs 0.8 pour k_1 et 1.8 pour k_2 . Ainsi, si on prend systématiquement la valeur $k=1.3$, milieu de l'intervalle, on sera toujours dans l'intervalle de comparaison $[0.8, 1.8]$ si on estime σ par s tel que $0.615 < \frac{s}{\sigma} < 1.385$, c'est-à-dire si on surestime ou si on sousestime σ de pas plus que 38.5 %.

Il nous arrivera parfois de vouloir comparer les EQMI de deux estimateurs particuliers. Pour cela, on calculera le rapport des deux EQMI.

Définition :

Soient deux familles d'estimateurs et $EQMI(T)$ et $EQMI(T')$ leur EQM asymptotique intégré respectif. On définit l'efficacité asymptotique de la première famille par rapport à la deuxième par :

$$e(T, T') = \frac{EQMI(T')}{EQMI(T)}$$

La contemplation des figures II.4.3.a1 - c22 montrant les EQM asymptotiques des familles d'estimateurs nous suggère que la comparaison des EQMI n'est pas toujours suffisante pour en déduire qu'une famille d'estimateurs est plus appropriée qu'une autre famille pour une ou plusieurs situations de contamination données. En effet, bien que $EQMI(T')$ puisse être supérieur à $EQMI(T)$, on pourrait préférer T'_n à T_n si la variabilité de $EQM[T]$ est plus grande que la variabilité de $EQM[T']$. La figure II.4.2.a illustre cette situation.

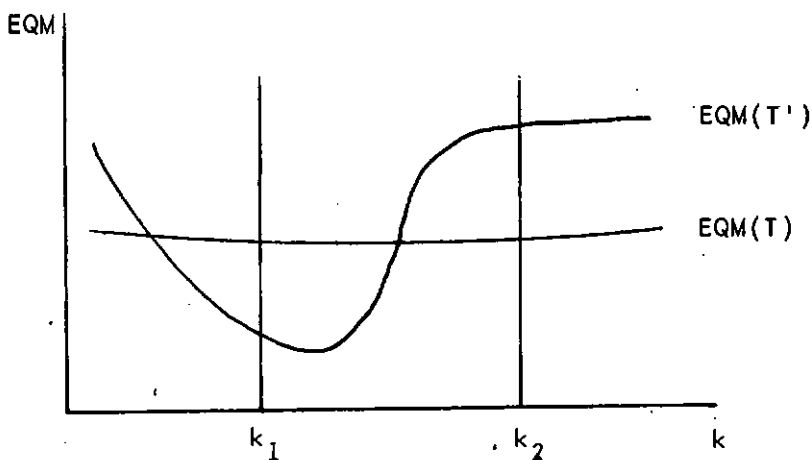


Figure II.4.2.a

Un bon choix de k permettra d'obtenir d'excellents résultats en prenant l'estimateur T_n . Par contre, un choix moins judicieux fournira des résultats bien inférieurs avec T_n qu'avec T'_n . Or il faut garder à l'esprit qu'en pratique le centre de symétrie aura une certaine variabilité due à l'estimation de σ . On préférera alors un estimateur moins efficace qu'un autre en moyenne, mais qui au moins ne sera jamais catastrophique. On appréciera donc d'avoir un estimateur relativement indépendant de k . Cette indépendance sera étudiée à travers la notion de stabilité. Un estimateur sera dit stable si son EQM asymptotique ne dépend que peu de k . La notion de stabilité n'est pas absolue, mais relative. Cela signifie que nous comparerons des estimateurs entre eux et en déduirons qu'un estimateur est plus ou moins stable qu'un autre estimateur. Dans ce but, nous construisons une courbe de stabilité ainsi définie :

Définition :

Soit l'ensemble des valeurs de $k \in [k_1, k_2]$ pour lesquelles l'EQM[T] est inférieur à une valeur fixée z , noté

$$K_T(z) = \{k \in [k_1, k_2] \mid \text{EQM}[T] < z\}$$

et soit $\mu_T(z)$ la mesure de Lebesgue pour $K_T(z)$. Alors on définit la stabilité de l'EQM[T] par :

$$S_T(z) = \frac{\mu_T(z)}{k_2 - k_1}$$

c'est-à-dire par la fraction de la longueur de l'intervalle $[k_1, k_2]$ qui se situe pour la courbe de l'EQM[T] sous la cote z .

Graphiquement, nous obtenons les courbes de la figure II.4.2.b. Celle-ci montre que $S_{T_1}(z)$ croît plus rapidement que $S_T(z)$. Cela signifie que l'EQM asymptotique de T'_n est plus stable que l'EQM asymptotique de T_n , pour $k_1 \leq k \leq k_2$.

Sans modifier la définition de $k_T(z)$, nous aurions pu définir la stabilité en fonction de u et non de z , en définissant u par :

$$u = z - \min_{k \in [k_1, k_2]} \text{EQM}[T] \quad , \quad \text{ou bien par}$$

$$u = \frac{z - \min_{k \in [k_1; k_2]} \text{EQM}[T]}{\min_{k \in [k_1; k_2]} \text{EQM}[T]}$$

Si nous y avons renoncé, c'est pour les raisons suivantes :

La première mesure supprimerait l'effet de grandeur du EQM asymptotique, ce qui se traduirait par le fait que les courbes $S_{(\cdot)}(z)$ auraient toutes la même origine, et cela provoquerait donc une perte d'information.

En plus de l'effet de grandeur, la deuxième mesure supprimerait l'effet d'amplitude par la réduction de l'échelle. Ainsi, des EQM ayant pour le même point k leur minimum, ceux-ci étant différents et ayant de plus des amplitudes différentes, pourraient être représentés par la même courbe de stabilité.

La courbe de stabilité peut être représentée de deux façons :

En termes de robustesse : un estimateur dont la courbe de stabilité croît rapidement peut être considéré comme un estimateur peu sensible au choix du paramètre k . C'est important, car en pratique il est impossible de choisir une valeur de k qui soit optimale, c'est-à-dire qui minimise l'EQM asymptotique de l'estimateur utilisé.

En termes probabilistes : en supposant que le k choisi

peut se situer indifféremment sur l'intervalle $[k_1, k_2]$, on peut interpréter k comme une variable aléatoire de distribution uniforme sur $[k_1, k_2]$. Alors $S_T(z)$ représente la probabilité que l'EQM de l'estimateur choisi appartenant à la famille T soit inférieure à z .

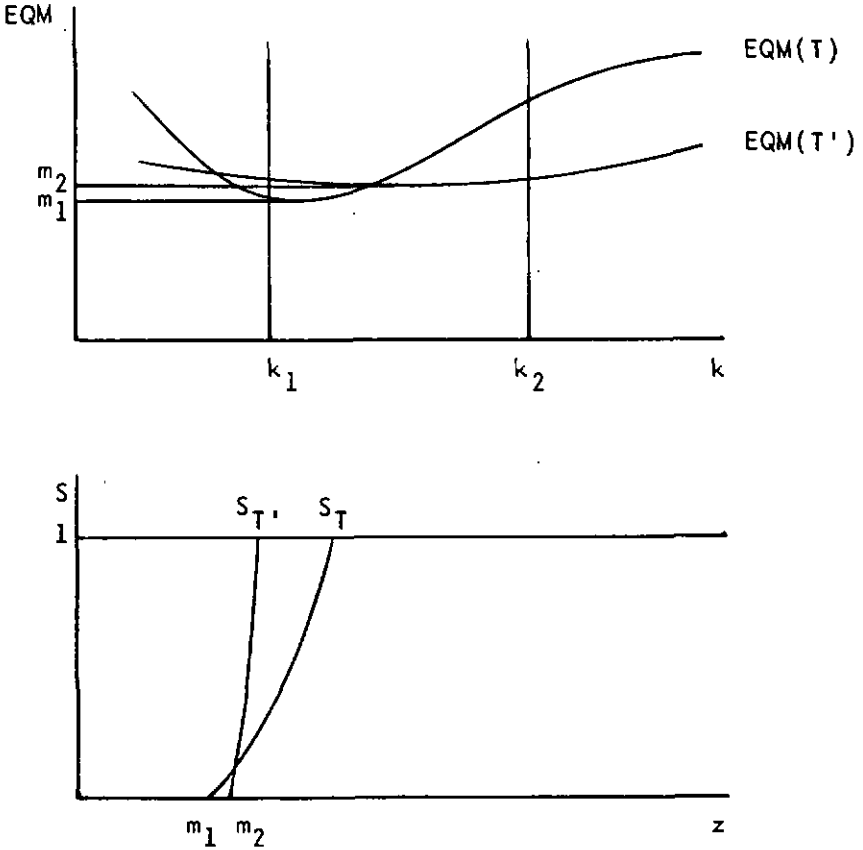


Figure II.4.2.b

II.4.3. Comparaisons des familles d'estimateurs

Nous avons calculé les EQM asymptotiques pour les dix-huit estimateurs décrits dans le tableau II.2.1. Le détail des calculs se trouve en annexe I. Nous avons dessiné les EQM asymptotiques des dix-huit familles, pour des valeurs de k comprises entre 0.1 et 5.1, sur les figures II.4.3.a1 à II.4.3.c22. Sur chaque dessin se trouvent les EQM asymptotiques des estimateurs d'une même classe, pour une contamination particulière.

Remarques

- 1) Pour la famille $S(1;0;1)$, seul représentant de sa classe, nous avons dessiné plusieurs contaminations sur le même graphe.
- 2) Afin d'obtenir une base de comparaison visuelle des EQM asymptotiques qui soit plus aisée, nous avons reproduit l'EQM de la famille $A(1;0;1)$ sur tous les graphes des familles du type asymétrique. Nous avons fait de même avec la famille $S(1;0;1)$ (représentant l'estimateur de Huber) pour le type symétrique.

Commentaires sur les EQM asymptotiques (figures II.4.3.a1 - II.4.3.a11)

Classe $A(\alpha;0;1)$ - La valeur $\alpha=1.5$ fournit la meilleure famille d'estimateurs de cette classe, lorsque la contamination est nulle et symétrique. La famille

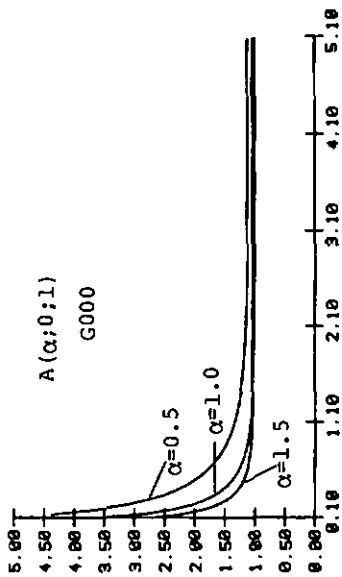


Figure II.4.3.a1

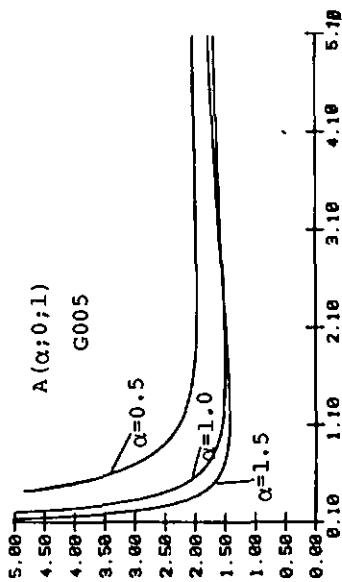


Figure II.4.3.a2

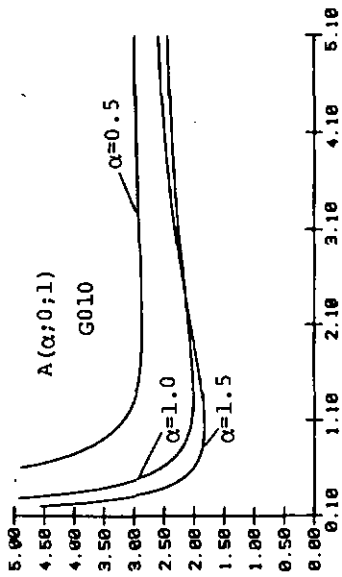


Figure II.4.3.a3

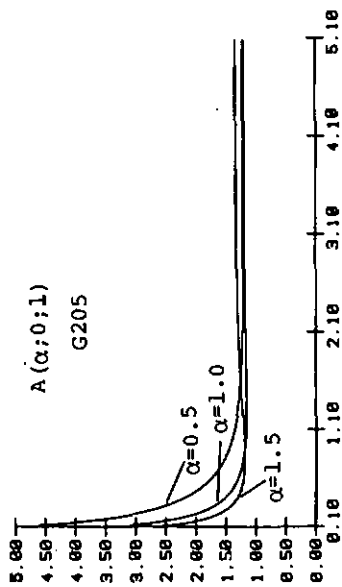


Figure II.4.3.a4

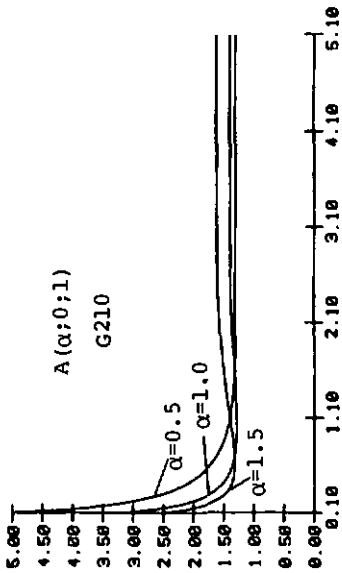


Figure II.4.3.05

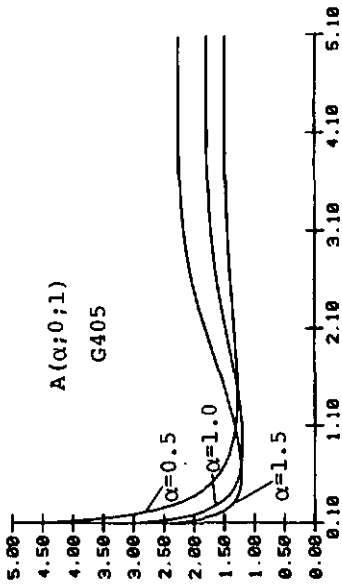


Figure II.4.3.06

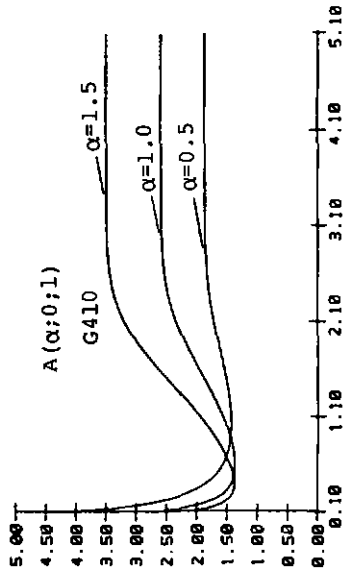


Figure II.4.3.07

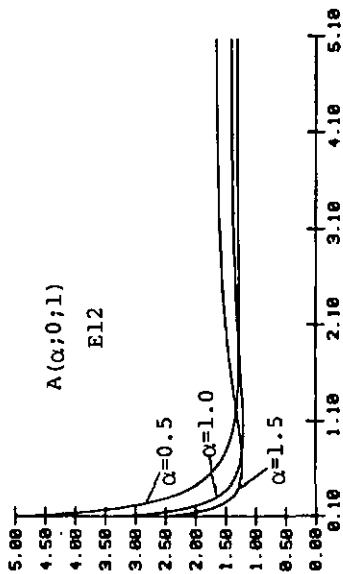


Figure II.4.3.a8

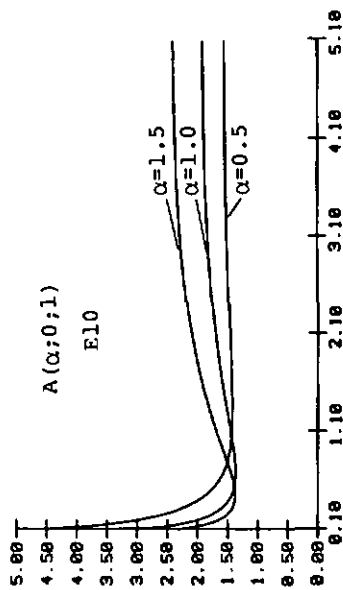


Figure II.4.3.a9

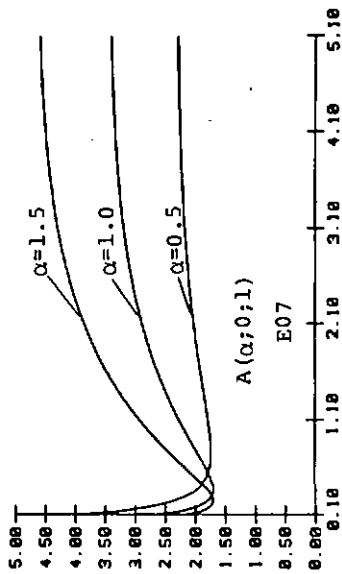


Figure II.4.3.a10

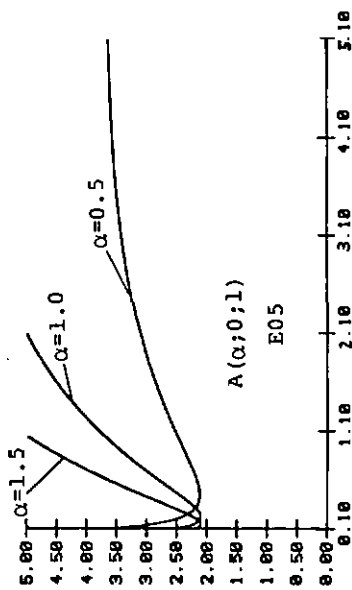


Figure II.4.3.a11

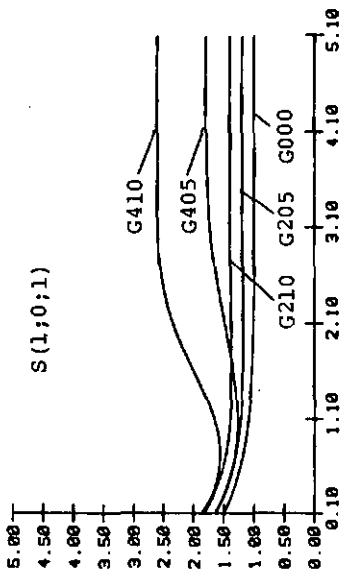


Figure II.4.3.a12

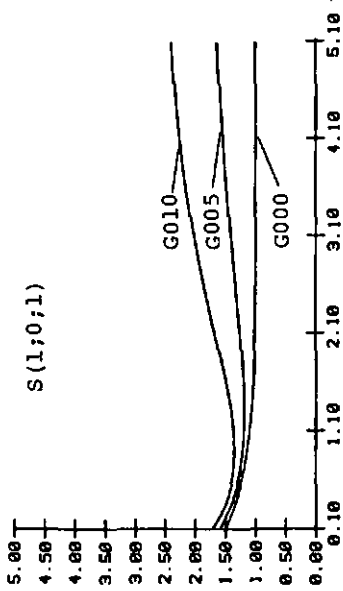


Figure II.4.3.a13

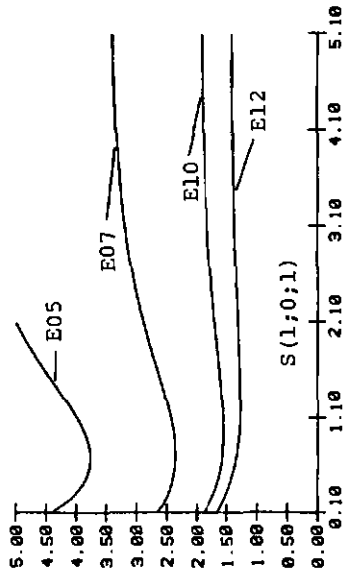


Figure II.4.3.a14

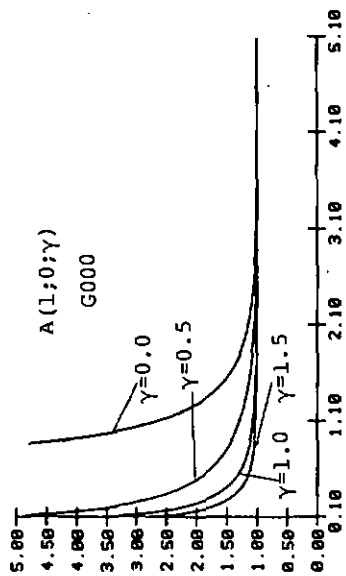


Figure II.4.3.b1

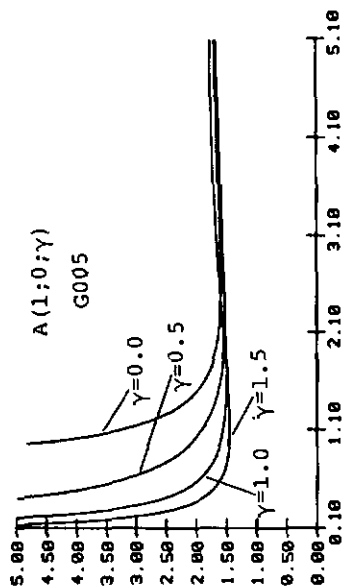


Figure II.4.3.b2

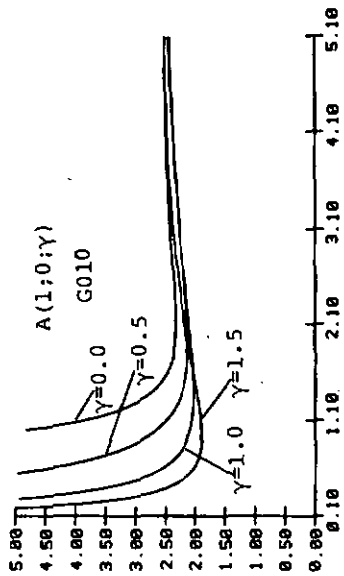


Figure II.4.3.b3

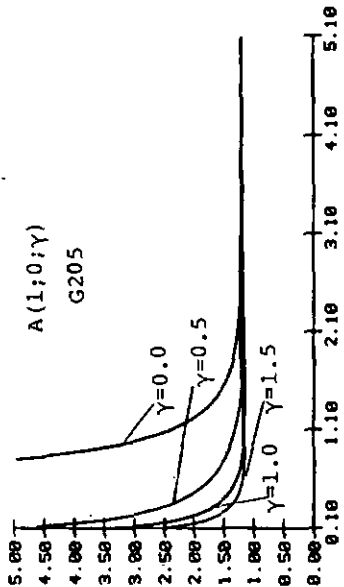


Figure II.4.3.b4

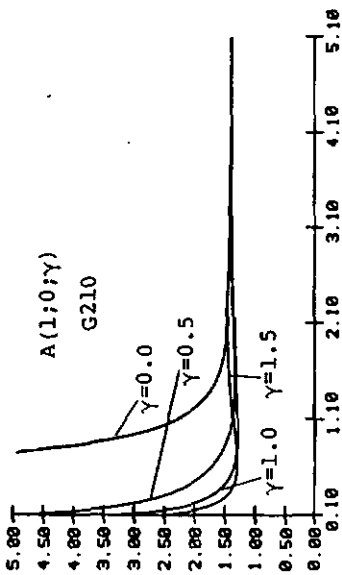


Figure II.4.3.b5

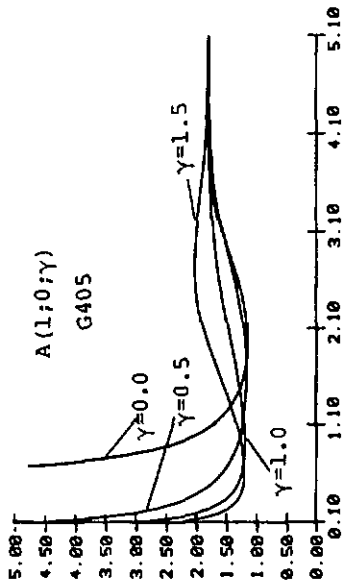


Figure II.4.3.b6

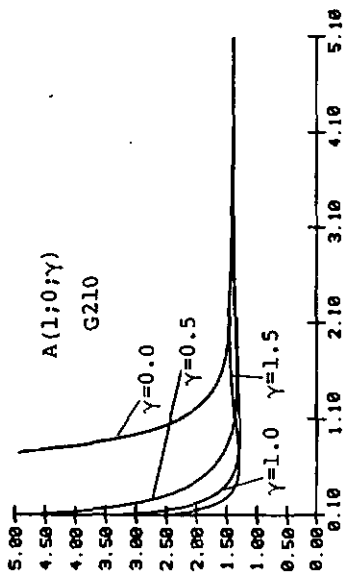


Figure II.4.3.b7

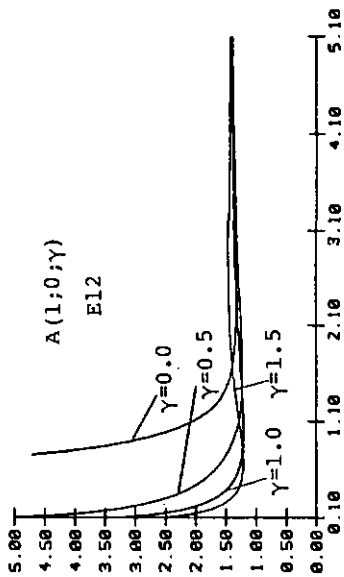


Figure II.4.3.b8

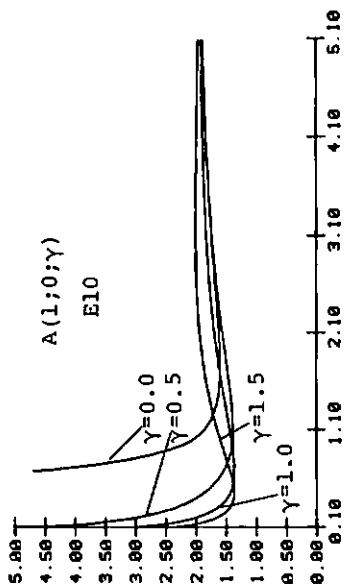


Figure II.4.3.b9

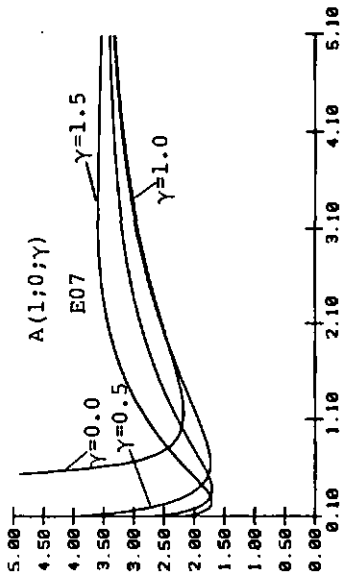


Figure II.4.3.b10

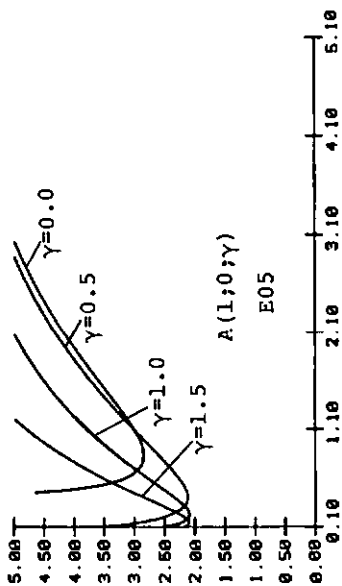


Figure II.4.3.b11

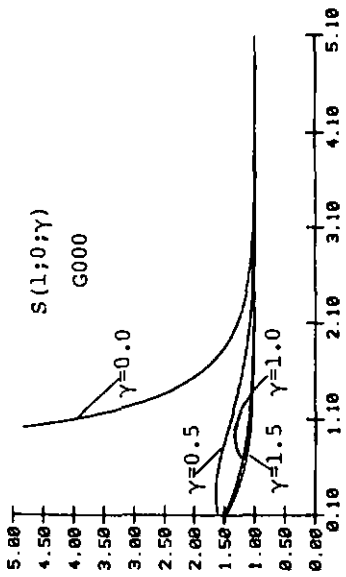


Figure II.4.3.bl2

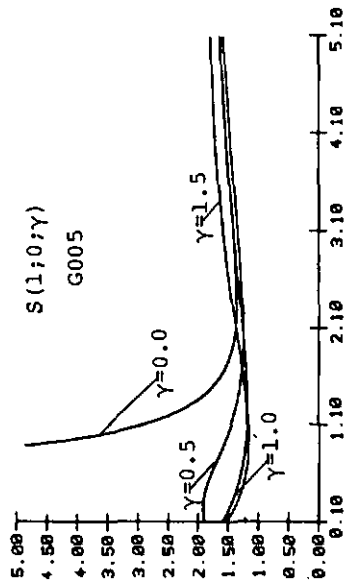


Figure II.4.3.bl3

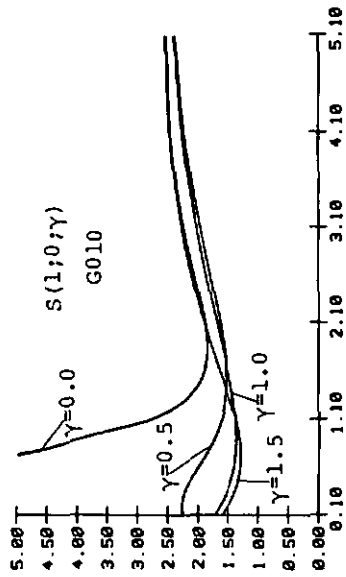


Figure II.4.3.bl4

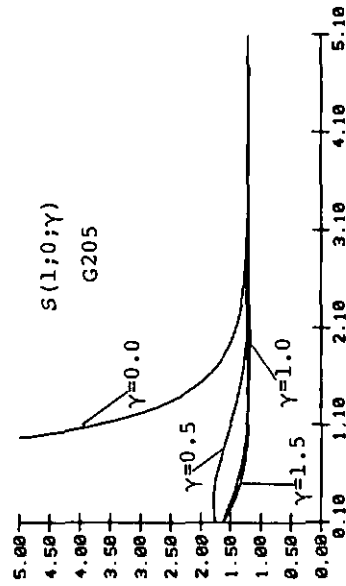


Figure II.4.3.bl5

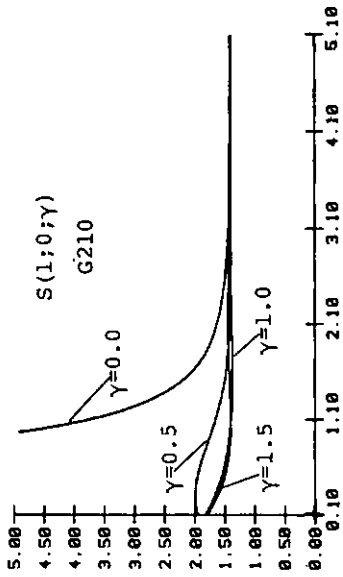


Figure II.4.3.bl6

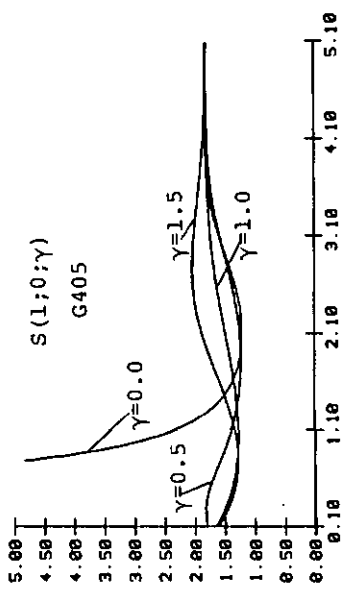


Figure II.4.3.bl7

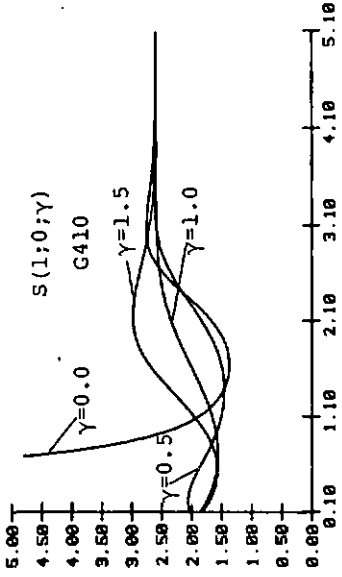


Figure II.4.3.bl8

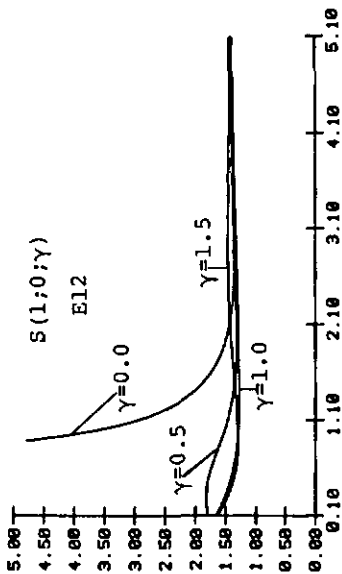


Figure II.4.3.b19

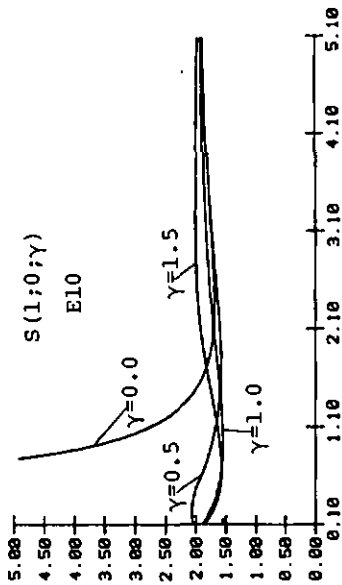


Figure II.4.3.b20

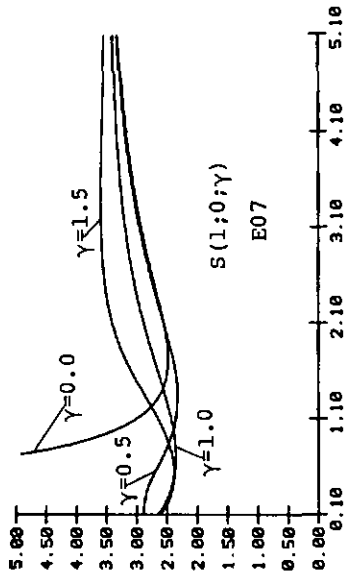


Figure II.4.3.b21

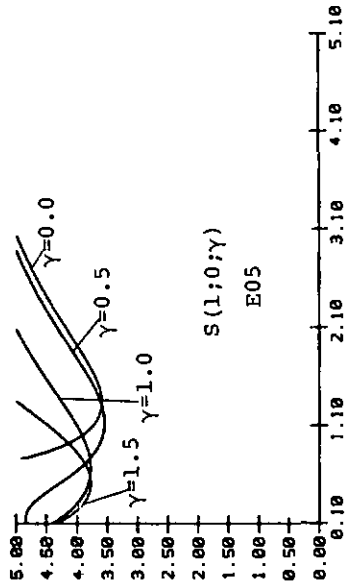


Figure II.4.3.b22

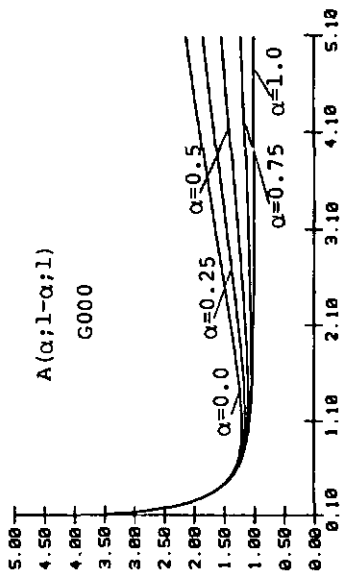


Figure II.4.3.c1

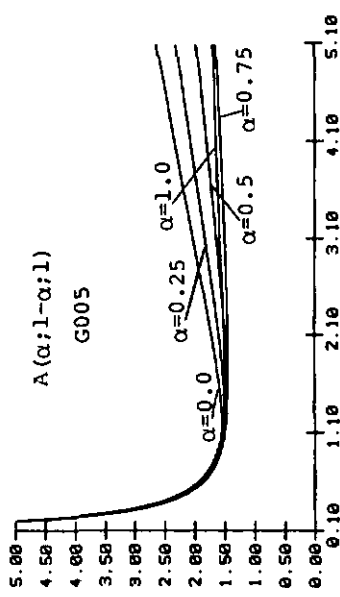


Figure II.4.3.c2

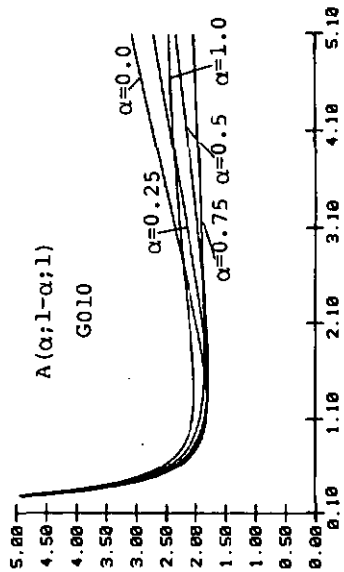


Figure II.4.3.c3

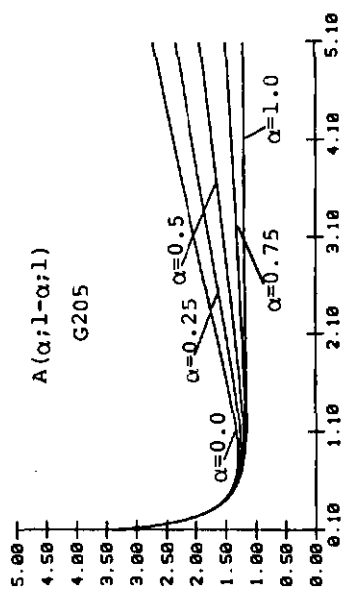


Figure II.4.3.c4

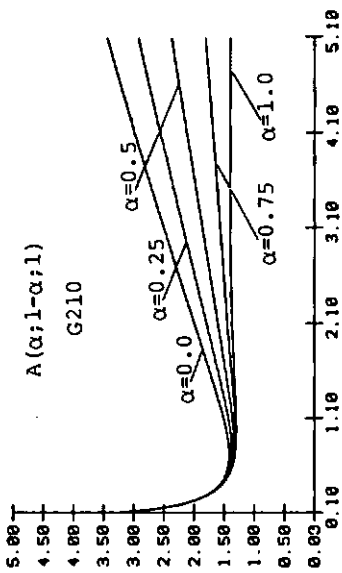


Figure II.4.3.c5

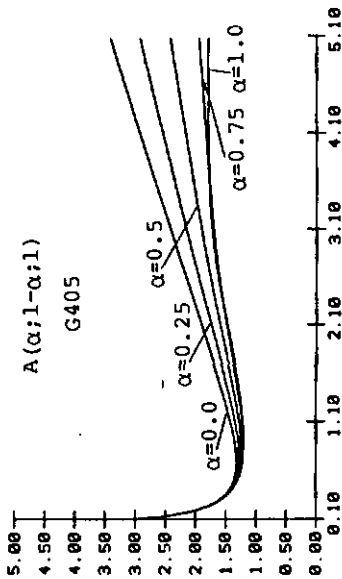


Figure II.4.3.c6

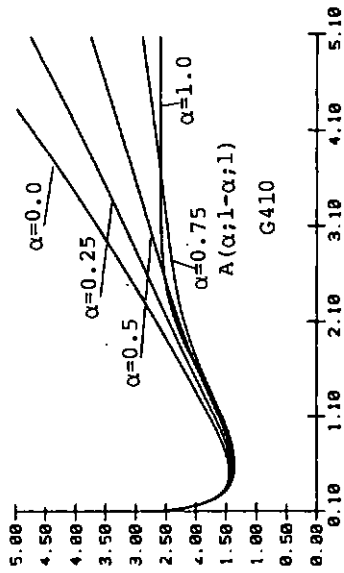


Figure II.4.3.c7

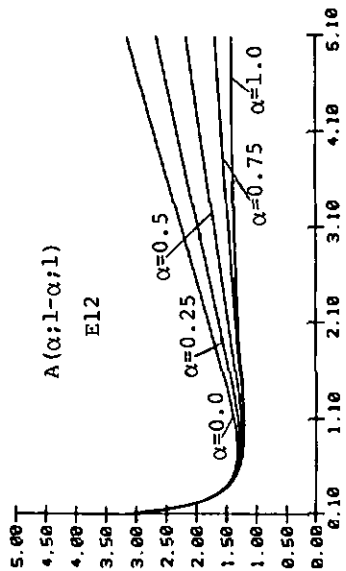


Figure II.4.3.c8

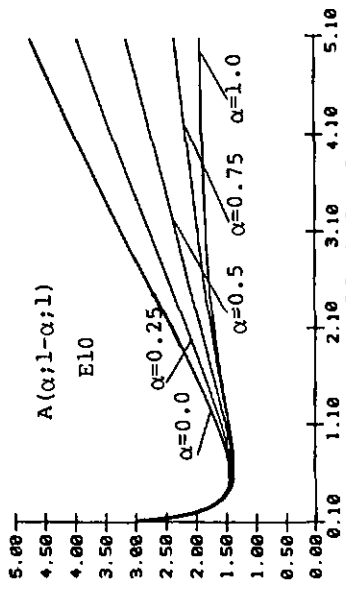


Figure II.4.3.c9

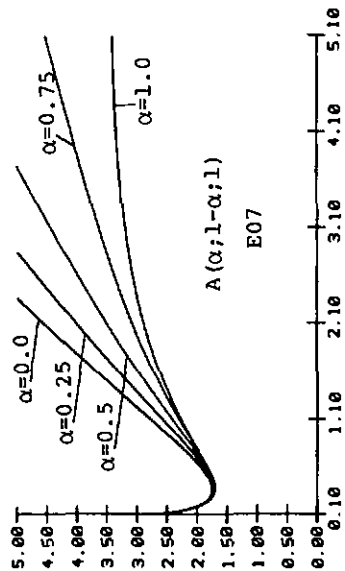


Figure II.4.3.c10

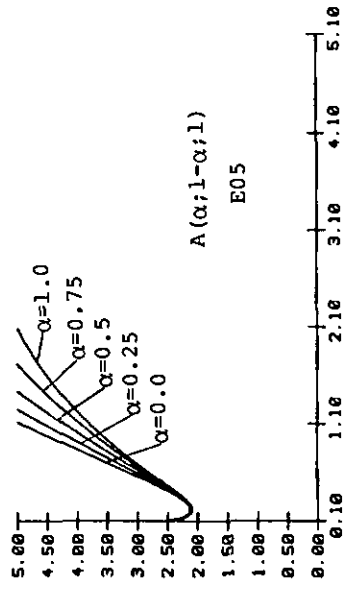
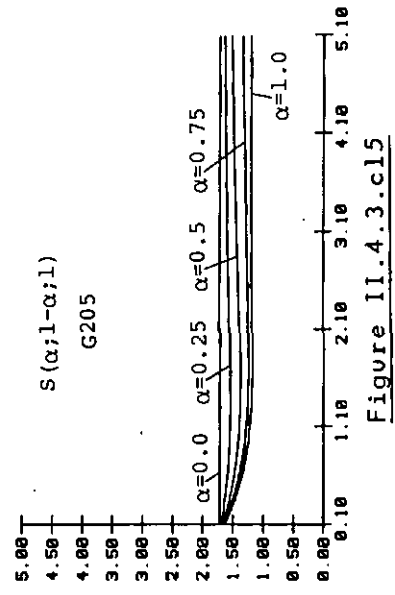
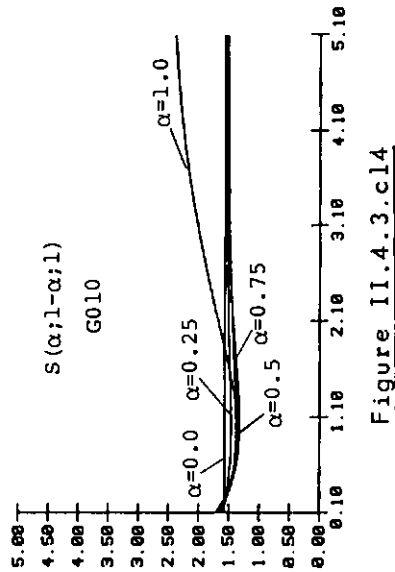
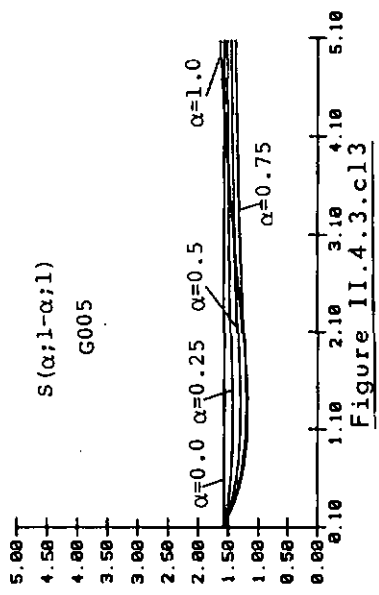
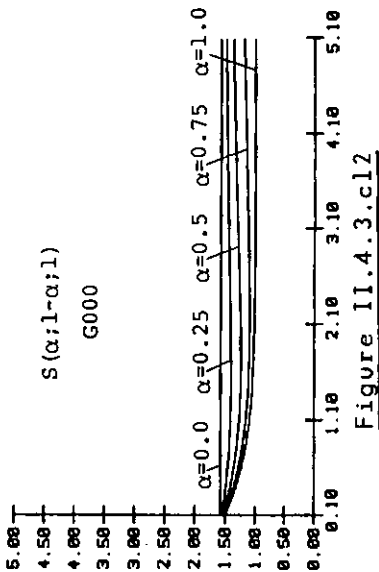


Figure II.4.3.c11



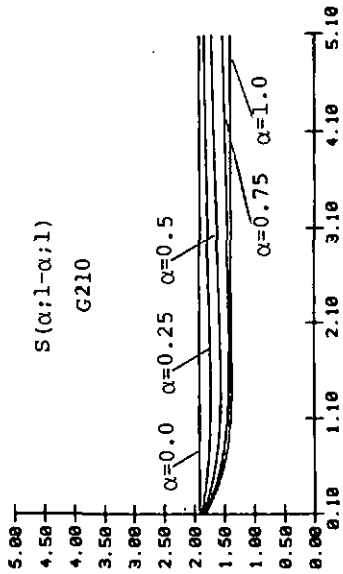


Figure II.4.3.cl6

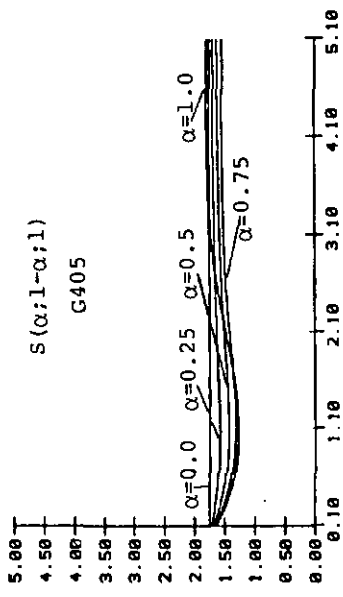


Figure II.4.3.cl7

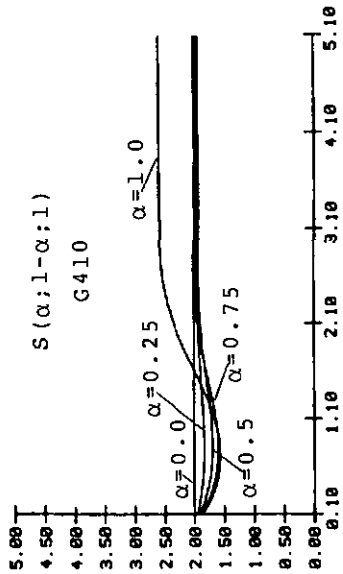


Figure II.4.3.cl8

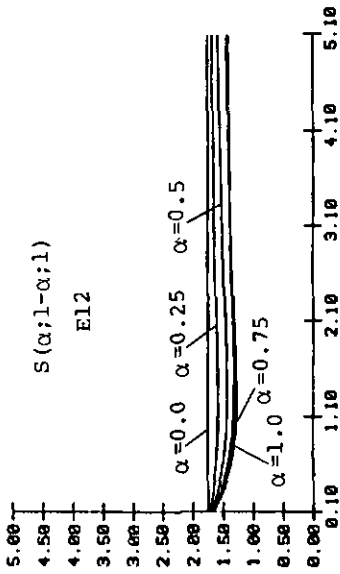


Figure II.4.3.c19

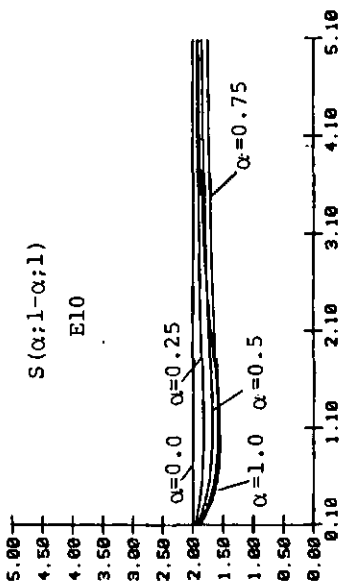


Figure II.4.3.c20

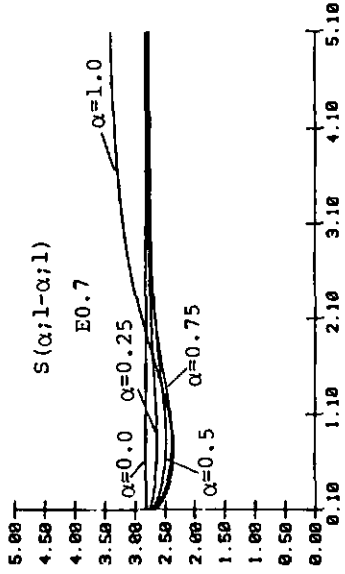


Figure II.4.3.c21

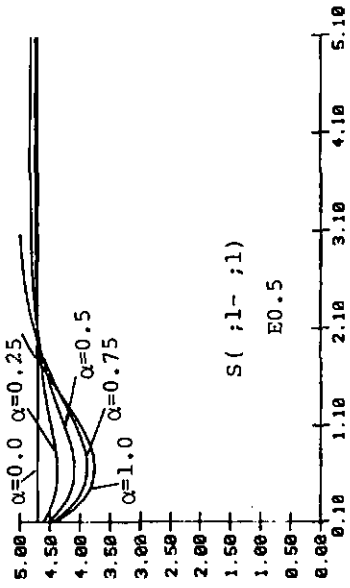


Figure II.4.3.c22

$\alpha=0.5$ est médiocre dans ces situations. La famille $\alpha=1.0$ se situe entre les deux précédentes mais son comportement est proche de la famille $\alpha=1.5$.

Dès que la contamination devient asymétrique, le comportement change totalement. Déjà pour G205 les EQM des familles se croisent entre 0.8 et 1.8. La famille $\alpha=1$ domine les deux autres familles. Toutefois on remarque que le comportement de la famille $\alpha=0.5$ s'améliore alors que celui de la famille $\alpha=1.5$ se dégrade. Cette tendance se confirme pour des contaminations plus prononcées. Pour E10, la famille $\alpha=1.5$ est complètement dominée par les deux autres; $\alpha=0.5$ est remarquablement stable; son EQM passe de 1.5 pour $k_1=0.8$ à 1.45 pour $k_2=1.8$ alors que l'EQM de la famille $\alpha=1.5$ passe, pour les mêmes bornes, de 1.5 à 2.0 et de 1.38 à 1.65 pour la famille $\alpha=1.0$. Il y a donc croisement des EQM des familles $\alpha=0.5$ et $\alpha=1.0$. On retrouve une situation analogue pour G405 mais un peu moins prononcée du fait que la contamination est plus faible.

Les figures des situations G410, E07 et E05 sont éloquentes. La seule famille acceptable pour cette classe est $\alpha=0.5$. Par exemple, pour E07, l'EQM passe de 1.73 à 1.98 entre k_1 et k_2 alors qu'il passe de 2.01 à 2.70 pour $\alpha=1.0$ et de 2.51 à 3.67 pour $\alpha=1.5$. Les différences de forme des EQM dans les situations G410 et E07 proviennent des formes des densités correspondantes. Pour G410, les EQM se stabilisent à partir de $k=3$, c'est-à-dire à partir de

$k\sigma_x = 4.68$, alors que pour E07 les EQM continuent d'augmenter à cause de la contamination exponentielle (e^{-x}) qui prolonge son action plus loin que la contamination gaussienne (e^{-x^2}).

Les points où les familles atteignent leur minimum fluctuent considérablement en fonction des situations et se déplacent vers la gauche à mesure que la contamination augmente. Si on élimine le cas G000 pour lequel l'EQM atteint son minimum lorsque k tend vers l'infini, on s'aperçoit que l'EQM minimum varie entre les valeurs suivantes de k :

0.55	et	2.40	pour	$\alpha = 0.5$
0.25	et	1.60	pour	$\alpha = 1.0$
0.15	et	1.30	pour	$\alpha = 1.5$

Classe $S(\alpha;0;1)$ (figures II.4.3.012 - II.4.3.014)

Cette classe n'est constituée que d'une famille proposée par Huber : $\alpha=1$. Dans toutes les situations de contamination asymétrique, son EQM est supérieur à celui de la famille $A(1;0;1)$ entre k_1 et k_2 . Cela est dû au fait que $A(1;0;1)$ réduit le biais, ce que ne peut pas faire $S(1;0;1)$ du fait de sa symétrie. Par contre, pour G005 et G010, l'EQM de cette famille est nettement inférieur à celui des familles de la classe $A(\alpha;0;1)$. Cela est dû au fait que ces familles introduisent un biais et que leurs fonctions ψ ne sont pas bornées pour les x négatifs, ce qui augmente la variance lorsqu'il y a contamination.

L'amplitude du EQM minimum pour cette famille (0.7 à 2.0) est plus faible que pour les familles de la classe $A(\alpha;0;1)$, ce qui constitue un avantage.

Classe $A(1;0;\gamma)$ (figures 11.4.3.b1 - 11.4.3.b11)

La famille $\gamma=0$ est très instable. Il ne faudrait jamais choisir pour une telle famille k inférieur à 1.5, car pour de petites valeurs de k , le biais devient très grand. Pour presque toutes les contaminations, cette famille n'offre pas d'intérêts (sauf dans les situations extrêmes G410 et E05).

La famille $\gamma=1.5$ est la meilleure de sa classe pour les contaminations symétriques (G005 et G010). Par contre elle perd toutes ses qualités en situation de contamination asymétrique. Elle est notamment toujours dominée par la famille $\gamma=1$ et, pour les moyennes et fortes contaminations (E10, E07, E05, G405 et G410) également par la famille $\gamma=0.5$.

Les familles intéressantes sont donc $\gamma=1.0$ (qu'on a déjà vu plus haut) et $\gamma=0.5$. La première famille est du point de vue de l'EQM meilleure que la seconde pour des contaminations faibles (G205 et G210). Pour des contaminations plus grandes, la seconde est préférable.

L'EQM minimum se situe entre :

0.85	et	4.00	pour	$\gamma = 0.0$
0.40	et	2.00	pour	$\gamma = 0.5$
0.25	et	1.60	pour	$\gamma = 1.0$
0.15	et	1.1	pour	$j = 1.5$

Classe $S(1;0;\gamma)$ (figures 11.4.3.b12 - 11.4.3.b22)

On retrouve la même instabilité de la famille $\gamma=0$ que dans la classe précédente. La forme des EQM des familles de cette classe est semblable à celle des EQM des familles de la classe précédente qui possèdent un EQM un peu inférieur pour des valeurs raisonnables de k , par réduction du biais, mais un EQM bien plus élevé pour des valeurs trap proches de zéro, à cause d'une augmentation sensible du biais. Remarquons que l'explosion du EQM de la famille $\gamma=0.0$ pour de petites valeurs de k n'est pas due au biais. Pour s'en convaincre, il suffit d'examiner le compartement de cette famille dans la situation G000 pour laquelle le biais de cette famille d'estimateurs est nul. Pour des trap petites valeurs de k , la formule de la variance asymptatique est peut-être fausse (voir la mise en garde de Huber (1964) concernant la validité des calculs de la variance asymptatique pour des fonctions ψ non monotones). Cette famille d'estimateurs également proposée par Huber (1964) ne devrait pas s'utiliser avec des valeurs de k inférieures à 1.5.

L'EQM minimum se situe entre :

1.35	et	3.80	pour	$\gamma = 0.0$
1.15	et	2.70	pour	$\gamma = 0.5$
0.70	et	2.00	pour	$\gamma = 1.0$
0.50	et	1.30	pour	$\gamma = 1.5$

Classe A($\alpha; 1-\alpha; 1$) (figures II.4.3.c1 - II.4.3.c11)

La famille $\alpha=0$ est toujours mauvaise, quelles que soient les contaminations. La famille $\alpha=1.0$ a un bon comportement pour toutes les contaminations asymétriques. Toutefois elle est plus instable que les autres familles dans la mesure où ses performances sont amoindries dans les cas de contamination symétrique, particulièrement G010. Les familles $\alpha=0.75$ et $\alpha=0.5$ sont les plus intéressantes de cette classe car elles allient les bonnes performances de la famille $\alpha=1$ en contamination asymétrique et la résistance de la famille $\alpha=0$ envers la forme de la contamination.

L'EQM minimum des familles d'estimateurs de cette classe se situe entre :

0.25	et	1.60	pour	$\alpha = 1.0$
0.25	et	1.80	pour	$\alpha = 0.75$
0.25	et	1.70	pour	$\alpha = 0.50$
0.25	et	1.30	pour	$\alpha = 0.25$
0.25	et	1.20	pour	$\alpha = 0.0$

Classe S($\alpha; 1-\alpha; 1$) (figures II.4.3.c12 - II.4.3.c22)

Cette classe d'estimateurs a été étudiée par

Dodge (1982). La famille $\alpha=1$ a un comportement très bon lorsque la contamination asymétrique et symétrique est faible (G205, G210, G405, E12, G005). Par contre, relativement aux autres familles de cette classe, cette famille est très sensible à l'intensité de la contamination : la différence de son EQM pour G005 et G010 est éloquent. La famille $\alpha=0$ est indépendante de k et représente la médiane. Il est intéressant de remarquer que cette famille a un EQM toujours supérieur à l'EQM des autres familles entre k_1 et k_2 . Bien que la médiane soit un estimateur connu comme très résistant à des valeurs aberrantes, il ne faudrait l'utiliser que dans des cas extrêmes, car son efficacité est médiocre et on peut généralement se servir d'autres estimateurs offrent une bonne protection face aux valeurs aberrantes et avec une efficacité plus élevée (par exemple la famille $\alpha=0.5$ ou $\alpha=0.75$).

L'EQM minimum des familles de cette classe se situe entre :

0.70	et	2.00	pour	$\alpha = 1.0$
0.70	et	1.80	pour	$\alpha = 0.75$
0.65	et	1.60	pour	$\alpha = 0.50$
0.60	et	1.50	pour	$\alpha = 0.25$

En résumé on peut dire que l'EQM des familles d'estimateurs du type $A(\alpha; \beta; \gamma)$ varie plus que celui des familles du type $S(\alpha; \beta; \gamma)$ pour des petites valeurs

de k . Ceci constitue un avantage pour les estimateurs du type symétrique puisque leur EQM asymptotique sera moins sensible au choix du paramètre k que l'EQM asymptotique des familles du type asymétrique. La médiane ($S(0;0;1)$) n'est jamais un estimateur intéressant pour les situations de contamination considérées. Même la moyenne arithmétique a un EQM inférieur à celui de la médiane pour les situations G000 (évidemment), G005, G205, G210, E12 et E10. De même la moyenne "coupée" $S(1;0;0)$ (à ne pas confondre avec la moyenne α -tronquée) et son correspondant asymétrique $A(1;0;0)$ donnent toujours de mauvais résultats.

Toutefois, mis à part pour quelques cas simples, la comparaison visuelle des EQM des familles d'estimateurs est difficile pour classer ces familles, car certaines sont performantes pour plusieurs contaminations et deviennent de piètres estimateurs dans d'autres cas de contamination. Par exemple la famille $A(0.5;0;1)$ se montre meilleure que son correspondant symétrique $S(0.5;0;1)$ uniquement pour des contaminations asymétriques moyennes et élevées.

Afin de pouvoir tirer des conclusions plus précises, nous avons dessiné sur les figures II.4.3.d1 à II.4.3.d11, les courbes de stabilité des familles d'estimateurs entre les bornes $k_1=0.8$ et $k_2=1.8$. Chaque figure représente la courbe de stabilité des dix-huit familles d'estimateurs pour une contamination particulière. La liste des noms des familles

G000

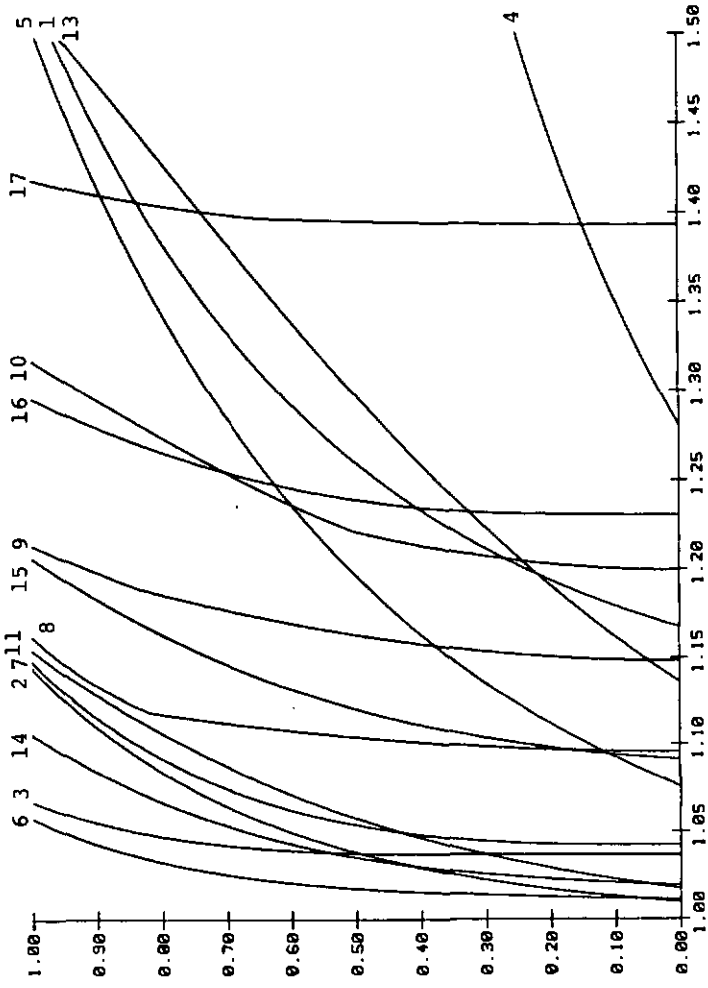


Figure II.4.3.d1

G005

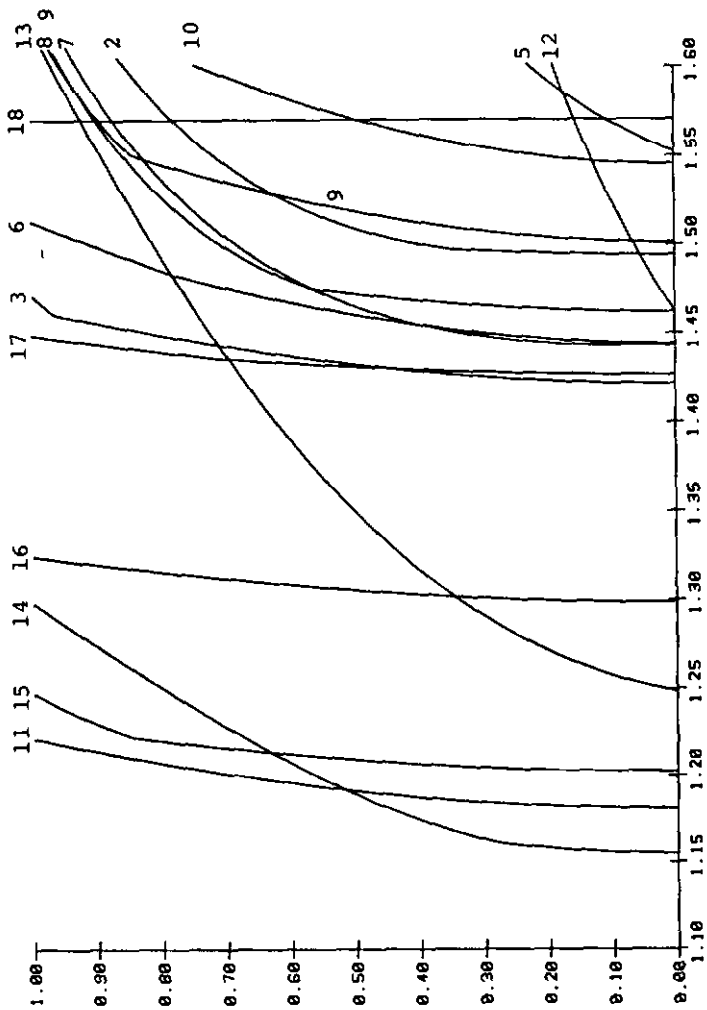


Figure II.4.3.d2

G010

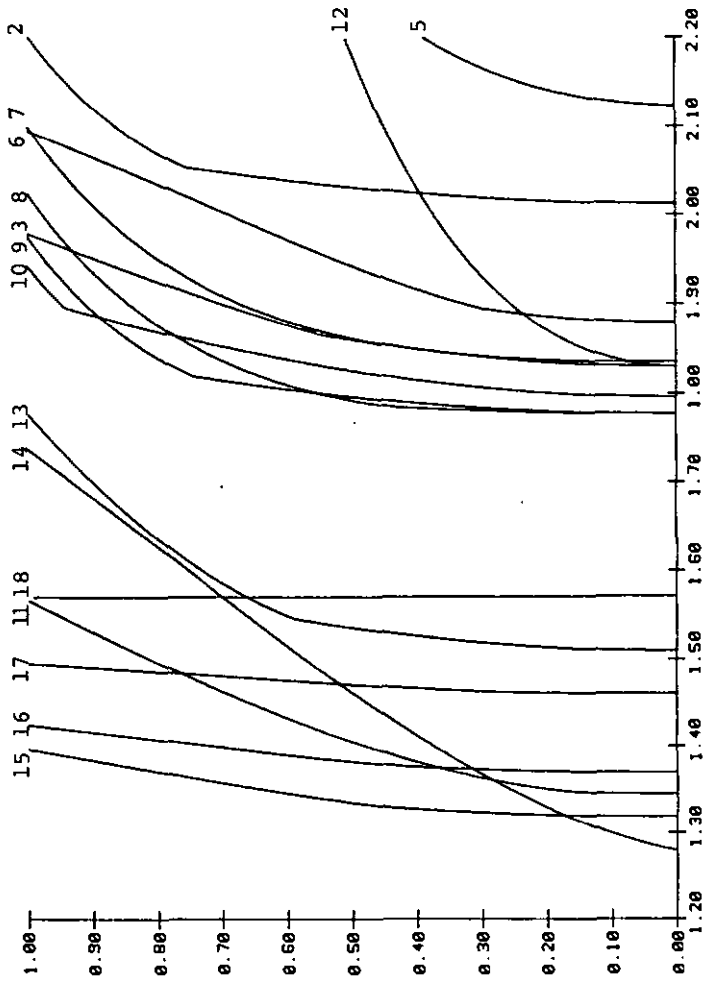


Figure II.4.3.d3

G205

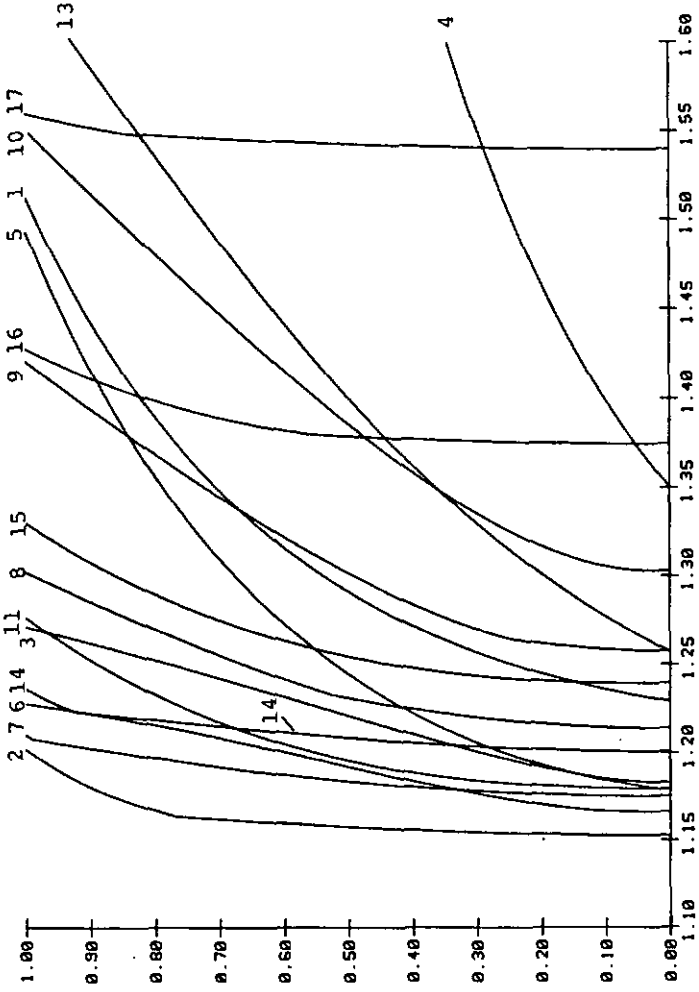


Figure II.4.3.d4

G210

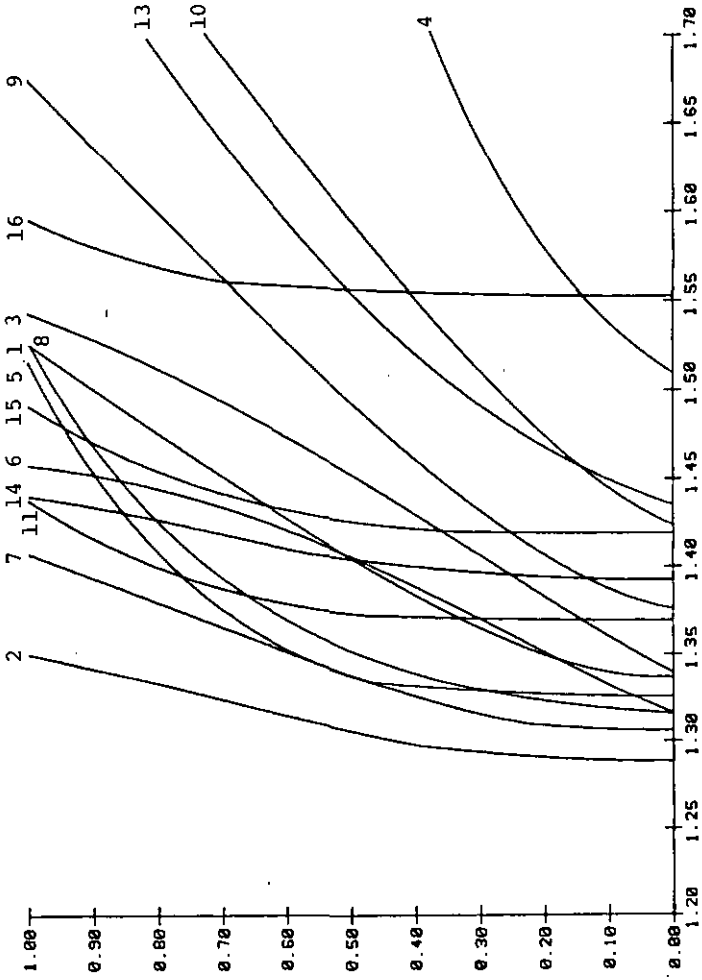


Figure II.4.3.d5

G405

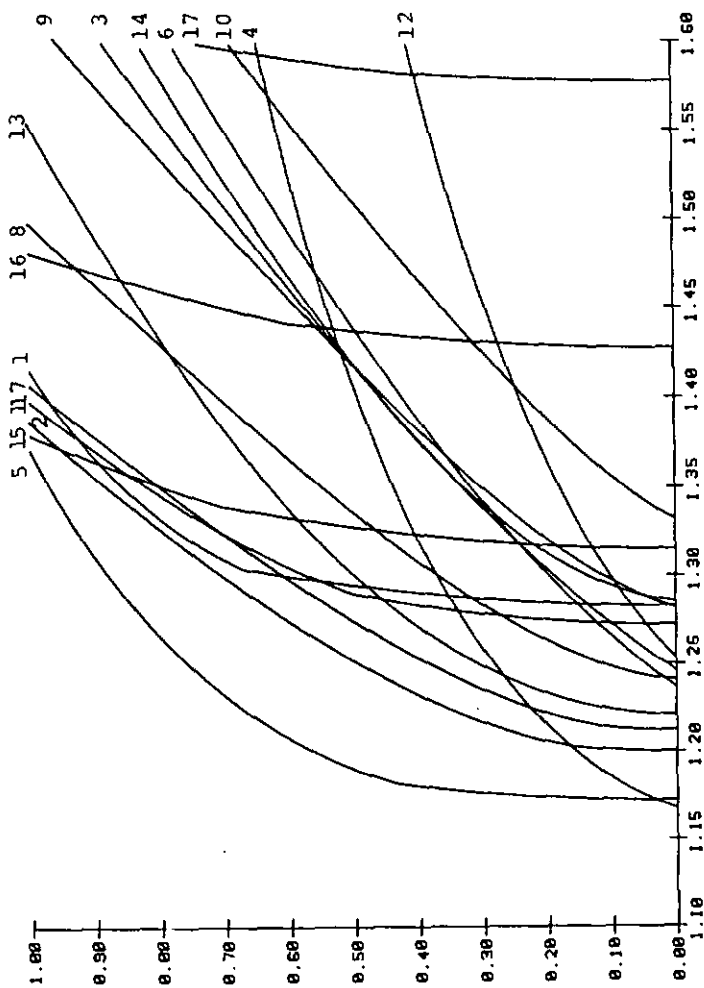


Figure II.4.3.d6

G410

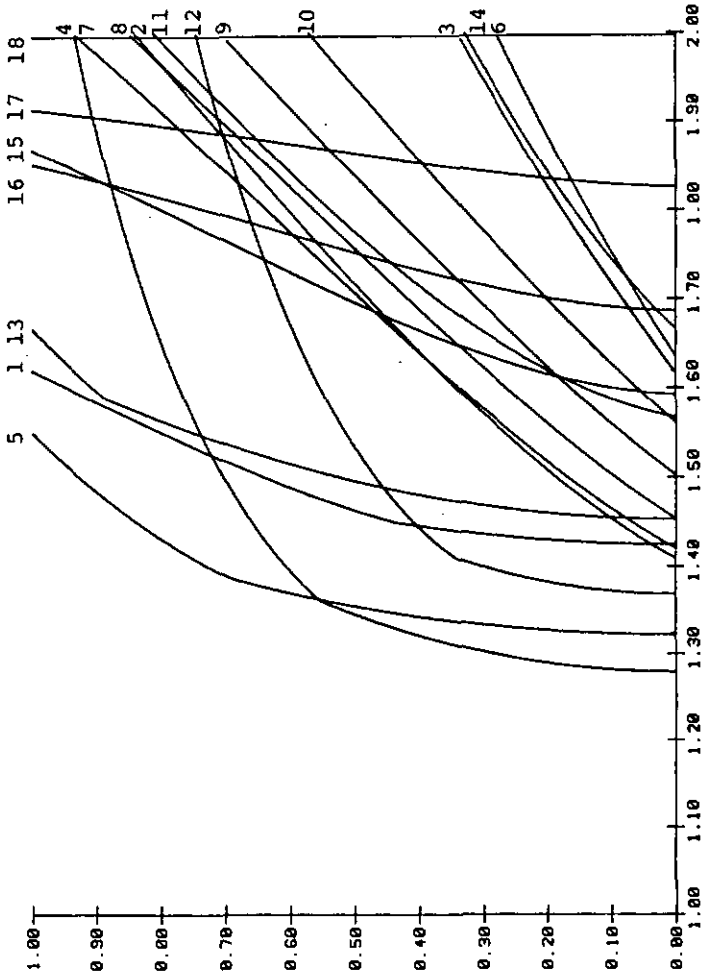


Figure II.4.3.d7

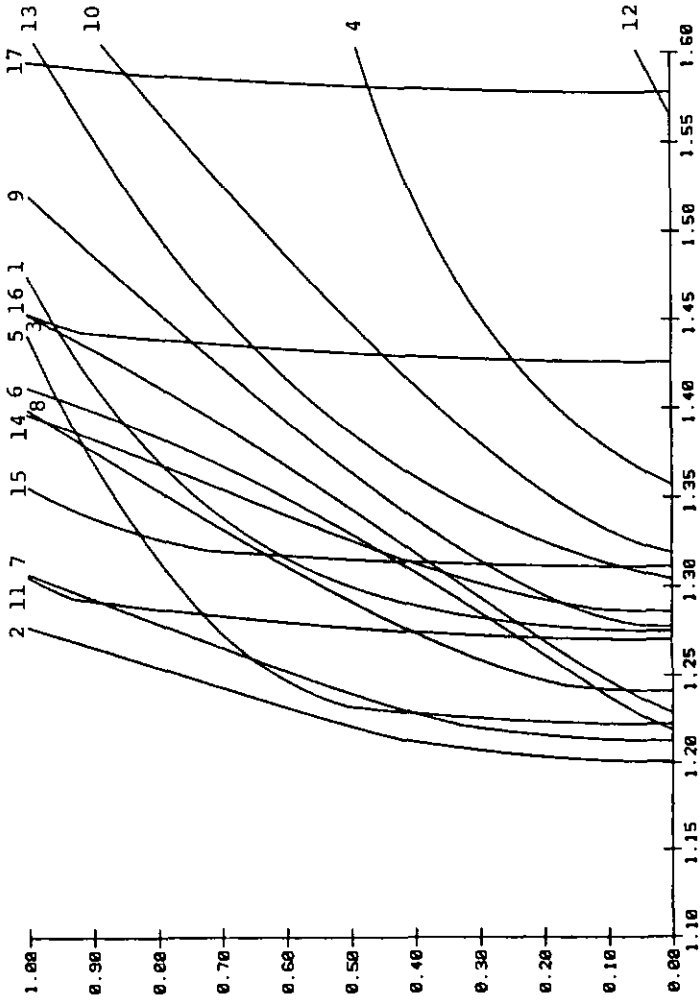


Figure II.4.3.d8

E10

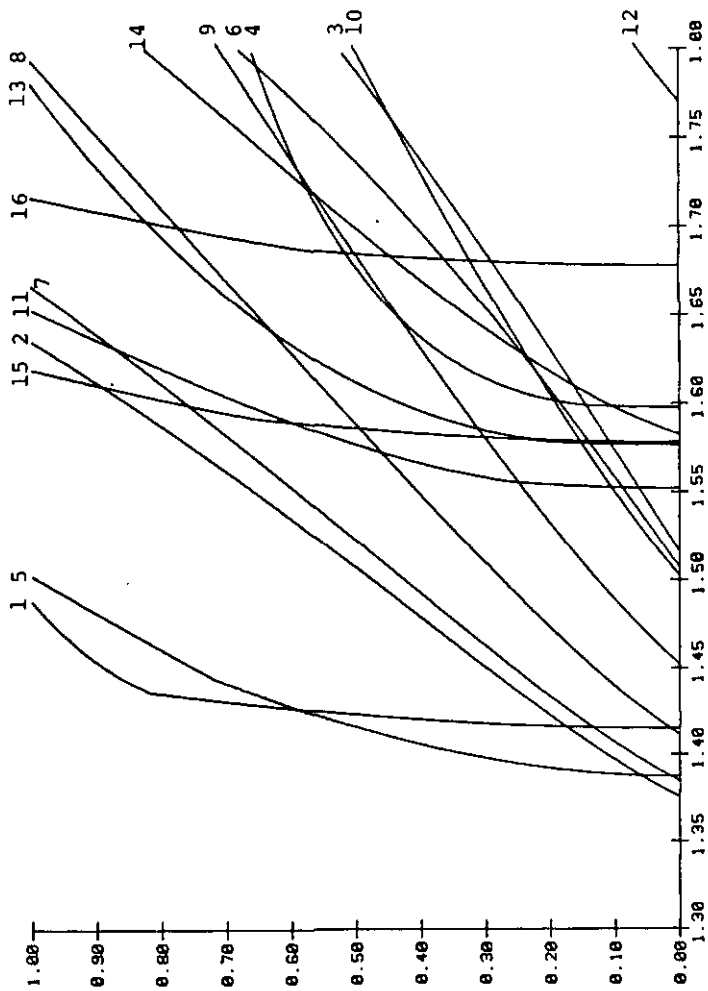


Figure II.4.3.d9

E07

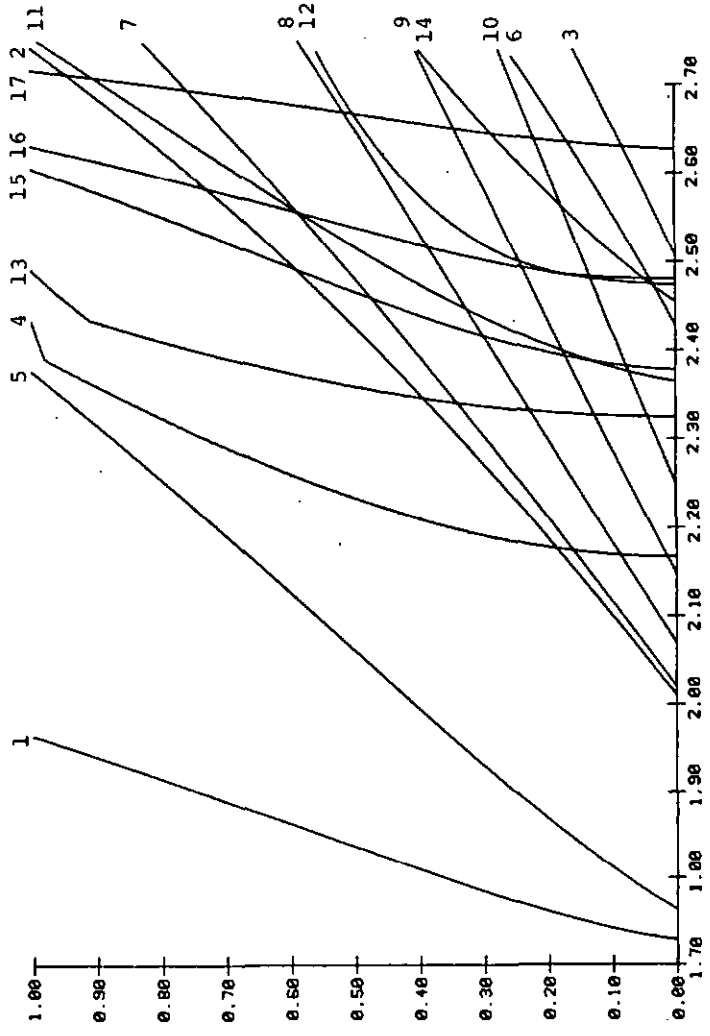


Figure II.4.3.d10

E05

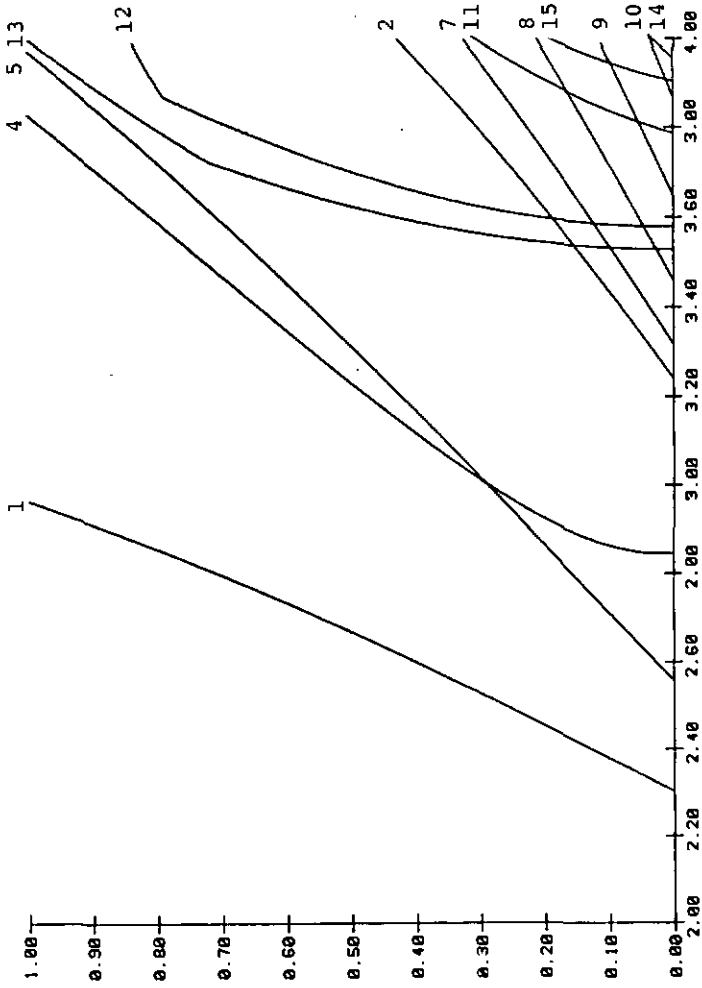


Figure II.4.3.d11

correspondants aux numéros attribués aux courbes de stabilité se trouve sur les tableaux II.4.2 et II.4.3. Sur ces figures, on remarque que certaines familles d'estimateurs sont complètement dominées par d'autres, car la courbe de stabilité de leur EQM se trouve toujours à droite de celle d'une autre famille. Dans d'autres cas, certaines courbes de stabilité se croisent et alors aucune famille ne domine une autre famille.

Commentaires sur les courbes de stabilité

D'une manière générale, les familles d'estimateurs du type $S(\alpha; \beta; \gamma)$ sont plus stables que les familles d'estimateurs du type $A(\alpha; \beta; \gamma)$. Cette constatation déjà faite auparavant se remarque au fait que la courbe de stabilité de leur EQM croît plus rapidement que celle du EQM des familles du type $A(\alpha; \beta; \gamma)$. Ces figures montrent à quel point les familles d'estimateurs asymétriques sont sensibles à la contamination. Alors que la famille $A(1; 0; 1)$ domine toutes les autres familles pour E_{12} par une courbe de stabilité remarquable, elle est presque complètement dominée par les familles $A(0.5; 0; 1)$ et $A(1; 0; 0.5)$ pour E_{10} dont les courbes de stabilité sont peu performantes pour E_{12} . On observe donc une inversion des performances de ces familles d'estimateurs pour des contaminations assez proches.

Par contre, les familles d'estimateurs symétriques conservent des courbes de stabilité très ana-

logues d'une situation à l'autre. Les familles les plus stables sont celles de la classe $S(\alpha; 1-\alpha; \gamma)$, $\alpha \neq 1$, autrement dit celles qui possèdent une discontinuité au point $x=c$.

Des courbes de stabilité on peut déduire les EQM minimum pour chaque contamination :

<u>Contamination</u>	<u>EQM minimum</u>	<u>Famille d'estimateurs</u>
G000	1.01	$S(1;0;1.5)$ et $A(1;0;1.5)$
G005	1.15	$S(1;0;1.5)$
G010	1.28	$S(1;0;1.5)$
G205	1.15	$A(1;0;1)$
G210	1.26	$A(1;0;1)$
G405	1.17	$A(1;0;0.5)$
G410	1.28	$A(1;0;0)$
E12	1.20	$A(1;0;1)$
E10	1.38	$A(1;0;1)$
E07	1.73	$A(0.5;0;1)$
E05	2.30	$A(0.5;0;1)$

Notons que l'EQM minimum pour les contaminations nulles et symétriques est atteint par la famille d'estimateurs qui se rapproche des procédures de windsorisation ($S(1;0;1.5)$). Toutefois l'EQM de cette famille est très instable comparé à celui des familles $S(0.75;0.25;1)$ ou même $S(1;0;1)$. Lorsque la contamination est nulle, on s'aperçoit que les meilleures familles d'estimateurs sont : $S(1;0;1.5)$, $A(1;0;1.5)$, $A(1;0;1)$, $A(1.5;0;1)$, $A(0.75;0.25;1)$ et $S(1;0;1)$. La

famille $S(1;0;1)$ est celle dont la variation du EQM est la plus grande (1.02 à 1.15) parmi les six familles du groupe de tête.

Les différences de stabilité entre familles s'accroissent avec l'augmentation de la contamination. Pour les contaminations asymétriques faibles et moyennes (G205, G210, E12), la famille dominante est $A(1;0;1)$. Lorsque la contamination asymétrique s'accroît (G405, G410, E10), cette famille est moins performante que la famille $A(1;0;0.5)$. Enfin lorsque la contamination asymétrique devient forte (E07 et E05), c'est la famille $A(0.5;0;1)$ qui domine toutes les autres.

Dans les tableaux II.4.2 et II.4.3, nous avons résumé les figures II.4.3 en attribuant des appréciations aux estimateurs, par situation de contamination. On interprètera les appréciations de la façon suivante :

*** : Cette famille d'estimateurs domine totalement ou presque totalement toutes les autres familles. Elle est la meilleure possible pour cette contamination.

Par domination presque totale on entend la chose suivante :

Soit T une famille d'estimateurs et $Y_T = \{y \mid S_T^{-1}(y) < S_{T'}^{-1}(y), T' \neq T\}$ et

μ_{Y_T} la mesure de Lebesgue de Y_T .
Alors T domine presque totalement
toutes les autres familles si
 $0.8 < \mu_{Y_T} < 1.0$.

** : Cette famille d'estimateurs est équivalente à une ou plusieurs autres familles d'estimateurs.

* : Cette famille d'estimateurs est admissible. Elle est totalement ou presque totalement dominée par une ou plusieurs autres familles. Cette famille est la moins bonne parmi celles du groupe de tête.

néant : Cette famille d'estimateurs n'est pas désirable. Elle est totalement dominée par une ou plusieurs autres familles.

Evidemment un problème de choix apparaît lorsque les courbes de stabilité de plusieurs familles se croisent, car on ne peut pas toutes les retenir. On préférera alors généralement la famille dont le EQMI est inférieur aux autres.

Les tableaux II.4.4 et II.4.5 donnent les valeurs des EQMI, les efficacités relatives par rapport à la famille dont l'EQMI est le plus petit, ainsi que

Familles d'estimateurs		Stabilité du EQM						
Numéros	Noms	G005	G010	G000	G205	G210	G405	G410
1)	A(0.5;0;1)					*	*	*
2)	A(1;0;1)			**	***	***	**	
3)	A(1.5;0;1)			**	*			
4)	A(1;0;0)							**
5)	A(1;0;0.5)					*	***	**
6)	A(1;0;1.5)			***	**	*		
7)	A(0.75;0.25;1)			*	**	**	*	
8)	A(0.5;0.5;1)							
9)	A(0.25;0.75;1)							
10)	A(0;1;1)							
11)	S(1;0;1)	**	**	*	*	*	*	
12)	S(1;0;0)							
13)	S(1;0;0.5)							*
14)	S(1;0;1.5)	**	*	**	*	*		
15)	S(0.75;0.25;1)	**	***					
16)	S(0.5;0.5;1)	*	**					
17)	S(0.25;0.75;1)		*					
18)	S(0;1;1)							

Tableau II.4.2

Familles d'estimateurs		Stabilité du EQM			
Numéros	Noms	E12	E10	E07	E05
1)	A(0.5;0;1)		**	***	***
2)	A(1;0;1)	***	*		
3)	A(1.5;0;1)				
4)	A(1;0;0)			*	**
5)	A(1;0;0.5)	*	**	**	**
6)	A(1;0;1.5)				
7)	A(0.75;0.25;1)	**	*		
8)	A(0.5;0.5;1)				
9)	A(0.25;0.75;1)				
10)	A(0;1;1)				
11)	S(1;0;1)	*			
12)	S(1;0;0)				
13)	S(1;0;0.5)			*	*
14)	S(1;0;1.5)				
15)	S(0.75;0.25;1)				
16)	S(0.5;0.5;1)				
17)	S(0.25;0.75;1)				
18)	S(0;1;1)				

Tableau II.4.3

Familles d'estimateurs	EQM asymptotique intégré et efficacité																				
	G005	G010	G000	G205	G210	G405	G410	G005	G010	G000	G205	G210	G405	G410							
A(0.5;0;1)	2.184	.544	16	3.096	.432	17	1.284	.793	13	1.311	.883	11	1.368	.954	4	1.303	.930	5	1.479	.928	3
A(1;0;1)	1.529	.778	12	2.040	.655	14	1.046	.973	4	1.157	1.000	1	1.305	1.000	1	1.259	.963	2	1.726	.795	7
A(1.5;0;1)	1.429	.832	7	1.869	.715	11	1.037	.982	2	1.216	.951	6	1.442	.905	10	1.423	.852	10	2.216	.620	16
A(1;0;0)	3.100	.384	18	3.776	.354	18	2.399	.424	17	2.186	.529	17	2.148	.608	17	1.594	.760	16	1.468	.935	2
A(1;0;0.5)	1.797	.662	15	2.378	.562	15	1.223	.832	10	1.269	.912	9	1.360	.960	3	1.212	1.000	1	1.372	1.000	1
A(1;0;1.5)	1.459	.815	8	1.949	.686	13	1.018	1.000	1	1.187	.975	3	1.392	.938	6	1.448	.837	13	2.305	.595	18
A(0.75;0.25;1)	1.485	.801	9	1.887	.709	12	1.062	.959	5	1.181	.978	2	1.338	.975	2	1.279	.948	3	1.706	.804	6
A(0.5;0.5;1)	1.489	.799	10	1.820	.735	9	1.105	.921	7	1.236	.936	7	1.406	.928	8	1.340	.904	8	1.754	.782	9
A(0.25;0.75;1)	1.521	.782	11	1.807	.740	8	1.163	.875	9	1.308	.885	10	1.496	.872	11	1.421	.853	9	1.838	.746	12
A(0;1;1)	1.571	.757	14	1.827	.732	10	1.231	.827	11	1.392	.831	13	1.598	.817	14	1.513	.801	14	1.940	.707	14
S(1;0;1)	1.189	1.000	1	1.415	.945	3	1.062	.959	5	1.201	.963	4	1.378	.947	5	1.301	.932	4	1.794	.765	11
S(1;0;0)	2.611	.456	17	2.616	.511	16	3.458	.294	18	3.253	.356	18	3.306	.395	18	2.086	.581	18	1.776	.773	10
S(1;0;0.5)	1.374	.865	5	1.565	.854	7	1.304	.781	14	1.412	.819	14	1.571	.831	13	1.326	.914	6	1.506	.911	4
S(1;0;1.5)	1.196	.994	2	1.469	.910	5	1.040	.979	3	1.205	.960	5	1.403	.930	7	1.438	.843	12	2.237	.613	17
S(0.75;0.25;1)	1.207	.985	3	1.337	1.000	1	1.125	.905	8	1.259	.919	8	1.430	.913	9	1.327	.913	7	1.699	.808	5
S(0.5;0.5;1)	1.301	.914	4	1.381	.968	2	1.242	.820	12	1.380	.838	12	1.554	.840	12	1.437	.843	11	1.749	.784	8
S(0.25;0.75;1)	1.426	.834	6	1.466	.912	4	1.393	.731	15	1.537	.753	15	1.719	.759	15	1.583	.766	15	1.857	.739	13
S(0;1;1)	1.563	.761	13	1.563	.855	6	1.563	.651	16	1.715	.675	16	1.906	.685	16	1.744	.695	17	1.987	.690	15

Tableau II.4.4

Familles d'estimateurs	EQM asymptotique intégré et efficacité											
	E12			E10			E07			E05		
A(0.5;0;1)	1.318	.927	8	1.422	.997	2	1.829	1000	1	2.644	1000	1
A(1;0;1)	1.222	1000	1	1.498	.947	3	2.398	.763	5	4.045	.654	6
A(1.5;0;1)	1.335	.915	10	1.769	.802	13	3.122	.586	17	5.503	.480	17
A(1;0;0)	1.912	.639	17	1.832	.774	16	2.239	.817	3	3.245	.815	2
A(1;0;0.5)	1.261	.969	3	1.418	1000	1	2.052	.891	2	3.281	.806	3
A(1;0;1.5)	1.317	.928	7	1.715	.827	12	2.919	.627	16	5.053	.523	15
A(0.75;0.25;1)	1.240	.985	2	1.515	.936	4	2.459	.744	6	4.263	.620	9
A(0.5;0.5;1)	1.294	.944	5	1.585	.895	7	2.631	.695	10	4.633	.570	12
A(0.25;0.25;1)	1.368	.893	11	1.684	.842	10	2.863	.639	14	5.082	.520	16
A(0;1;1)	1.454	.840	14	1.801	.787	14	3.128	.585	18	5.575	.474	18
S(1;0;1)	1.273	.960	4	1.578	.899	5	2.519	.726	8	4.194	.630	7
S(1;0;0)	2.671	.458	18	2.529	.561	18	2.879	.635	15	3.776	.700	5
S(1;0;0.5)	1.405	.870	12	1.628	.871	8	2.360	.775	4	3.653	.724	4
S(1;0;1.5)	1.324	.923	9	1.697	.836	11	2.822	.646	13	4.838	.547	14
S(0.75;0.25;1)	1.314	.930	6	1.583	.896	6	2.463	.743	7	4.220	.627	8
S(0.5;0.5;1)	1.426	.857	13	1.681	.844	9	2.530	.723	9	4.357	.607	10
S(0.25;0.75;1)	1.576	.775	15	1.824	.777	15	2.656	.689	11	4.522	.585	11
S(0;1;1)	1.743	.701	16	1.987	.714	17	2.804	.652	12	4.680	.565	13

Tableau II.4.5

le rang des familles d'estimateurs étudiées.

Commentaires sur les EQMI

En général on s'aperçoit de la concordance entre les appréciations données aux courbes de stabilité et le classement résultant des EQMI.

Lorsque la contamination est nulle, beaucoup de familles d'estimateurs conservent un EQMI faible : huit familles ont une efficacité relative supérieure à 0.90 et les familles les plus efficaces appartiennent au type asymétrique. Toutefois, c'est pour cette situation que l'efficacité relative est la plus étendue : de 0.294 pour la plus mauvaise famille ($S(1;0;0)$) à 1.00 pour la meilleure ($A(1;0;1.5)$).

Lorsque la contamination est symétrique, les meilleures familles d'estimateurs sont symétriques. La meilleure famille asymétrique pour G005 est $A(1.5;0;1)$ qui a une efficacité relative de 0.832, et pour G010, c'est la famille $A(0.25;0.75;1)$ dont l'efficacité relative est de 0.74.

Lorsque la contamination est asymétrique, les trois meilleures familles sont toujours asymétriques. La meilleure famille d'estimateurs symétriques est pour chaque contamination :

<u>Cantamination</u>	<u>famille</u>	<u>rang</u>	<u>efficacité relative</u>
G205	S(1;0;1)	4	0.963
G210	S(1;0;1)	5	0.947
G405	S(1;0;1)	4	0.932
G410	S(1;0;0.5)	4	0.911
E12	S(1;0;1)	4	0.960
E10	S(1;0;1)	5	0.899
E07	S(1;0;0.5)	4	0.775
E05	S(1;0;0.5)	4	0.724

Il n'y a que pour les contaminations E07 et E05 que l'efficacité relative de la meilleure famille symétrique descend nettement au-dessous de 0.9 . Toutefois, les familles asymétriques placées en rang 2 et 3 pour les mêmes contaminations ont également une efficacité relative inférieure à 0.9 .

II.4.4. Conclusions

- 1) Aucune famille asymétrique ne résiste à une contamination symétrique. Ce résultat est attendu car les fonctions ψ de ces familles d'estimateurs ne sont pas bornées à gauche et sont donc incapables de diminuer l'influence d'une contamination se situant à gauche du paramètre à estimer.
- 2) De même, les familles symétriques résistent très mal aux contaminations asymétriques. Les seules familles acceptables sont A(1;0;1), qui est robuste aux contaminations faibles et moyennes, et

$A(1;0;0.5)$, qui est robuste aux contaminations asymétriques fortes.

- 3) La famille la plus intéressante pour les fortes contaminations asymétriques est $A(0.5;0;1)$. Son EQMI reste sensiblement inférieur aux autres. Cette famille reste acceptable lorsque la contamination est moyenne. Par contre, elle est indésirable pour des contaminations symétriques ou nulles.
- 4) Le comportement de la famille $A(1;0;1)$ est bon lorsque la contamination asymétrique est faible et moyenne. Par contre, son EQMI augmente de manière sensible avec la contamination. Lorsque la contamination est symétrique ou nulle, cette famille est également indésirable. Les familles $A(0.5;0;1)$ et $A(1;0;1)$ sont donc complémentaires pour des contaminations asymétriques.
- 5) On retrouve la famille $A(1;0;0.5)$ parmi les meilleures familles dans toutes les situations de contamination asymétrique, sauf pour G205.
- 6) Lorsque la contamination est symétrique ou nulle, les deux seuls choix raisonnables sont $S(1;0;1)$ et $S(0.75;0.25;1)$.
- 7) Les familles les plus stables relativement au choix de k sont les familles symétriques ayant une discontinuité au point $x=c$.

- 8) La forme de contamination (exponentielle ou gaussienne) n'influence pas le comportement des estimateurs. De E12, contamination intermédiaire entre G205 et G210, résulte un classement des familles également intermédiaire aux deux classements obtenus avec G205 et G210.
- 9) Dans le cas de la contamination gaussienne, la plupart des familles sont sensibles à une combinaison des paramètres de contamination μ et ϵ et non à l'un ou à l'autre.

Le choix final d'une famille d'estimateurs est difficile. En simplifiant quelque peu, on peut opposer deux stratégies :

- 1) Choisir une famille particulière pour chaque contamination. Cette stratégie est utopique, car pratiquement on ne sait pas avec exactitude comment sont contaminées les données, ni même parfois si elles sont contaminées ou non. En adoptant une telle stratégie, on aboutirait souvent à des résultats plus mauvais que ceux obtenus par des méthodes classiques ne faisant pas de distinction entre bonnes données et données aberrantes. Une telle stratégie est toutefois possible, mais avec un autre type d'estimateur : les estimateurs adaptatifs.
- 2) Choisir une famille qui soit satisfaisante pour un éventail le plus large possible de contamination. Les estimateurs minmax sont issus d'une telle

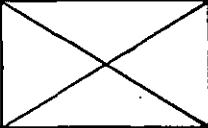
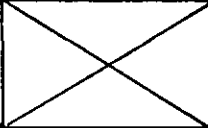
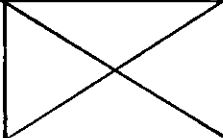
stratégie. Cette stratégie est celle qui s'avère la plus utile en pratique.

Dans notre étude, nous avons remarqué qu'aucune famille n'est capable de couvrir avec satisfaction les contaminations symétriques et asymétriques. On ne peut donc pas se contenter d'un seul estimateur pour ces deux types de contamination. Afin de faciliter le choix d'estimateurs, nous allons réduire la réalité à quelques situations types vraisemblables.

Divisons l'intensité de la contamination en deux classes : contamination faible d'une part et forte d'autre part. L'intensité peut être calculée comme une combinaison des paramètres de contamination. On peut par exemple prendre le biais de la moyenne ($\epsilon \cdot \mu$ pour le modèle de contamination gaussien). Nous faisons alors les hypothèses suivantes :

- 1) Si la contamination est faible, on ne peut pas déterminer si elle est symétrique ou asymétrique. On dira qu'elle est indéterminée.
- 2) Si la contamination est forte, on arrive à déterminer si elle est symétrique ou asymétrique.

Nous pouvons alors établir le tableau suivant qui résume le choix des estimateurs que l'on propose :

type de contam. intensité de contam.	symétrique	asymétrique	indéterminée
faible			$S(1;0;1)$ $k=1.4$
forte	$S(0.75;0.25;1)$ $k=1.25$	$A(0.5;0;1)$ $k=1.5$	

Lorsque la contamination est faible et supposée indéterminée, notre choix s'est finalement porté sur l'estimateur $S(1;0;1)$, $k=1.4$. Ce choix tient aux raisons suivantes :

- Bien que selon le tableau 11.4.2. les familles $S(1;0;1)$ et $S(1;0;1.5)$ paraissent plus ou moins équivalentes, l'efficacité relative de la première famille est toujours plus élevée que celle de la seconde, sauf pour G000.
- La courbe de stabilité de la famille $S(1;0;1.5)$ croît généralement moins vite que celle de la famille $S(1;0;1)$ et se dégrade rapidement lorsque la contamination devient forte (G405, G410). De plus cette famille est plus sensible au modèle de contamination (gaussien - exponentiel) que la famille $S(1;0;1)$.
- La famille $S(1;0;1)$ est représentée par une fonction

ψ continue, alors que la fonction ψ de la famille $S(1;0;1.5)$ présente une discontinuité au point $x=k\sigma+c$. La conséquence est que la sensibilité au voisinage du point de discontinuité des estimateurs de cette famille est plus grande que celle d'estimateurs dont la fonction ψ est continue.

Lorsque la contamination est forte et symétrique, nous avons choisi l'estimateur $S(0.75;0.25;1)$, $k=1.25$. Cet estimateur qui représente une combinaison convexe des normes L_1 et L_2 pour x compris entre $c - k$ et $c + k$ est remarquablement stable, comme d'ailleurs toutes les familles de la classe $S(\alpha;1-\alpha;1)$ $0 \leq \alpha < 1$ et qui sont d'autant plus stables que α se rapproche de 0. On préfère cette famille à la famille $S(1;0;1)$ bien qu'elle soit dominée par cette dernière pour G205 car sa stabilité est plus grande et plus indépendante de la forme de contamination. Toutefois, cette famille a le désavantage d'être discontinue au point $x=c$.

Enfin, lorsque la contamination est fortement asymétrique, la meilleure famille est $A(0.5;0;1)$. La valeur proposée pour k est 1.5. Cette valeur peut paraître petite. Elle a été obtenue en considérant les situations G405, G410, E12, E10 et E07. L'EQM de cet estimateur est optimum pour G405 et vaut :

1.52	pour	G410
1.30	pour	E12
1.44	pour	E10
1.89	pour	E07

II.5. Choix d'un estimateur robuste de dispersion

Puisque les M- estimateurs étudiés ne sont pas en général invariants d'échelle, il est nécessaire de trouver un estimateur de σ afin de pouvoir estimer la valeur de k dans l'échelle des données, à savoir $k\sigma$. Pour que l'estimation de $k\sigma$ ne soit pas sujette à de trop grandes fluctuations dues à la contamination, nous voulons que l'estimation de σ soit robuste. Le but est donc de trouver un estimateur de σ dont l'EQM asymptotique ne dépend que peu de la contamination. Nous voulons également que le biais de l'estimateur soit nul lorsqu'il n'y a pas de contamination. Comme cette étude ne porte pas spécialement sur les estimateurs de paramètres de dispersion, nous n'avons pas poussé la recherche très loin et nous n'avons retenu que trois estimateurs qui seront comparés entre eux :

- S_n^1 : la distance interquartile
- S_n^2 : l'écart-type empirique
- S_n^3 : l'écart absolu moyen

Nous donnons ici les expressions analytiques de l'espérance mathématique asymptotique et de la variance asymptotique de ces trois estimateurs. Le développement des calculs pour les contaminations considérées se trouve dans l'annexe II.

1) La distance interquartile

Cet estimateur est défini par :

$$S_n^1 = \frac{1}{2} \left(X_{\frac{3n}{4}} - X_{\frac{n}{4}} \right)$$

dont les propriétés asymptotiques sont :

$$\lim_{n \rightarrow \infty} E(S_n^1) = \frac{1}{2} \left(\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}} \right)$$

$$\lim_{n \rightarrow \infty} V(\sqrt{n} S_n^1) = \frac{1}{64} \left(\frac{3}{f^2(\xi_{\frac{1}{4}})} + \frac{3}{f^2(\xi_{\frac{3}{4}})} - \frac{2}{f(\xi_{\frac{1}{4}})f(\xi_{\frac{3}{4}})} \right)$$

où f est la densité de la loi considérée. (voir par exemple Kendall et Stuart (1, 1969), p. 239).

La contrainte s'exprime par le fait que si la contamination est nulle (f est une densité de Gauss de moyenne μ et variance σ^2), le biais est nul; soit

$$\xi_{\frac{3}{4}} = 0.6745 \cdot \sigma + \mu$$

$$\xi_{\frac{1}{4}} = 0.6745 \cdot \sigma + \mu$$

$$\frac{1}{2 \cdot 0.6745} \left(\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}} \right) = \sigma$$

Donc $\frac{1}{2 \cdot 0.6745} \left(X_{\frac{3n}{4}} - X_{\frac{n}{4}} \right)$ est un estimateur sans

biais de σ d'une loi normale.

2) L'écart-type empirique

Il est défini par :

$$S_n^2 = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

dont les propriétés asymptotiques sont :

$$\lim_{n \rightarrow \infty} E(S_n^2) = \sqrt{\mu_2} \quad , \text{ où } \mu_2 \text{ est le moment}$$

centré sur μ , d'ordre 2.

$$\lim_{n \rightarrow \infty} V(\sqrt{n} S_n^2) = \frac{\mu_4 - \mu_2^2}{4\mu_2}$$

A nouveau, si la contamination est nulle, le biais doit être nul. Or

$$\lim_{n \rightarrow \infty} E(S_n^2) = \sigma \quad \text{est sans biais.}$$

3) L'écart absolu moyen

Il est défini par :

$$S_n^3 = \frac{1}{n} \sum |X_i - X|$$

Pour une loi quelconque (μ, σ^2) les propriétés asymptotiques de cet estimateur sont (voir par exemple Kendall et Stuart (I, 1969) :

$$\lim_{n \rightarrow \infty} E(S_n^3) = E|X - \mu| \quad \text{et}$$

$$\lim_{n \rightarrow \infty} V(\sqrt{n} S_n^3) = \sigma^2 - (E|X - \mu|)^2$$

En particulier, si la contamination est nulle :

$$\lim_{n \rightarrow \infty} E\left(\sqrt{\frac{\pi}{2}} S_n^3\right) = \sigma \quad \text{est sans biais}$$

et sa variance asymptotique vaut $\sigma^2\left(\frac{\pi}{2} - 1\right)$.

Nous avons calculé les EQM asymptotiques de ces trois estimateurs de l'écart-type des lois considérées. Le tableau II.5.1. présente les résultats obtenus.

Commentaires

Dans tous les cas où la contamination est asymétrique unimodale, l'EQM asymptotique de S_n^3 est plus petit que celui de S_n^1 . Par contre, S_n^1 est meilleur en EQM que S_n^3 lorsque la contamination est symétrique et lorsque la loi devient bimodale.

L'EQM asymptotique de S_n^2 est plus faible que celui de S_n^1 et S_n^3 lorsqu'il n'y a pas de contamination (résultat attendu) et lorsque la contamination asymétrique est faible (G205 et G210). Par contre S_n^2 est très instable selon le modèle de contamination. En effet, bien que G210 et E12 soient proches, l'EQM de S_n^2 est très différent dans ces deux cas, alors que l'EQM de S_n^1 et S_n^3 ne sont guère différents dans ces deux situations. De plus, l'EQM de S_n^2 augmente rapidement avec la contamination. Donc S_n^2 n'est jamais à conseiller, sauf naturellement lorsqu'on est sûr qu'il n'y a pas de contamination.

On remarque également que les biais des estimateurs S_n^1 et S_n^3 sont toujours inférieurs à celui de S_n^2 . Alors que le biais de S_n^1 est toujours positif, celui de S_n^3 est négatif pour la plupart des contaminations asymétriques. Dans ces situations, S_n^3 sous-estime donc plus fortement σ que S_n^1 .

On peut donc conclure que si la contamination paraît être symétrique S_n^1 est plus approprié que S_n^3 . Par contre, si la contamination est asymétrique unimodale, il vaut mieux utiliser l'estimateur S_n^3 .

Tableau 11.5.1.

Chaque cellule contient : - le biais asymptotique
- l'EQM asymptotique

Estimateurs Contaminations	S_n^1	S_n^2	S_n^3
G000	0.0007 1.3694	0.0000 0.5000	0.0000 0.5708
G005	0.0749 1.5801	0.3229 5.5596	0.0382 1.6724
G010	0.0749 1.6968	0.5811 7.6627	0.0765 2.7740
G205	0.0378 1.5498	0.0909 0.7174	-0.0400 0.9492
G210	0.1119 1.8608	0.1662 0.8294	-0.0918 1.3198
G405	0.0749 1.7178	0.3266 2.2217	-0.0477 1.8600
G410	0.1490 2.3184	0.5620 2.6218	-0.1290 3.0907
E12	0.0749 1.7192	0.1816 1.5562	-0.0917 1.3767
E10	0.1490 2.3161	0.3567 2.1228	-0.0929 2.0771
E07	0.4085 4.2525	0.7564 5.0473	0.0312 3.7836
E05	0.7791 7.4802	1.2862 11.8587	1.2557 4.6989

Deuxième partie

APPLICATION AUX MODELES AUTOREGRESSIFS

Nous allons à présent nous pencher sur l'étude des modèles autorégressifs du premier ordre (AR(1)). L'intérêt de tels processus réside dans le fait que dans bien des domaines d'étude et en particulier en économie les séries chronologiques peuvent être souvent perçues comme des réalisations de tels processus.

Le premier chapitre de cette partie sera consacré à l'étude asymptotique des estimateurs. On étudiera d'abord les méthodes classiques en présentant leurs caractéristiques du point de vue de la robustesse. Ensuite, après avoir défini des modèles de contamination, on passera à l'étude des M-estimateurs des paramètres des modèles AR(1).

Le deuxième chapitre sera consacré à l'étude des estimateurs fondés sur des échantillons de taille finie par simulation et dans le troisième chapitre, nous traiterons un exemple d'une série chronologique contaminée.

Chapitre I. ETUDE ASYMPTOTIQUE

I.1. Approche classique des modèles autorégressifs

I.1.1. Processus stochastiques

Sait $X(e)$ une v.a., c'est-à-dire une fonction qui associe à chaque état e appartenant à un ensemble E une valeur notée x . Un processus stochastique est une fonction $X(e, t)$ qui associe à chaque état e une valeur notée $x(t)$ ou x_t pour t défini sur un ensemble θ . Si θ est défini sur l'ensemble Z , ce que nous supposerons dès maintenant, on dit que le processus stochastique est discret.

On peut considérer une série temporelle de n termes comme une réalisation d'un processus stochastique discret, c'est-à-dire comme une réalisation d'une suite de n v.a.

Afin de pouvoir déduire le comportement du processus au temps $t+1$ à partir du temps t , c'est-à-dire connaître la loi conditionnelle de X_{t+1} , il est nécessaire de connaître la loi de probabilité conjointe $p_{1,2,\dots,t+1}(X_1, X_2, \dots, X_{t+1})$. On peut alors en déduire la loi de probabilité conditionnelle de la v.a. X_{t+1} $p_{t+1|1,2,\dots,t}(X_{t+1}|X_1, X_2, \dots, X_t)$. Pour la prévision, cette démarche est malaisée et peu économe, puisque, à chaque époque t , il est nécessaire de connaître la loi de probabilité conjointe $p_{1,2,\dots,t}(X_1,$

X_2, \dots, X_t).

Afin d'éviter cette difficulté on va spécifier un mécanisme ou modèle qui liera la suite $\{X_t; t \in \mathbb{Z}\}$ des v.a.

1.1.2. Processus stochastiques discrets stationnaires

On dit qu'un processus stochastique discret est strictement stationnaire si toutes ses propriétés statistiques sont invariantes par une translation temporelle, c'est-à-dire si

$$p(X_t, X_{t+1}, \dots, X_{t+k}) = p(X_{t+m}, X_{t+m+1}, \dots, X_{t+m+k})$$

quels que soient t , m et k appartenant à \mathbb{Z} . On en déduit que les probabilités marginales sont identiques : $p(X_t) = p(X_{t+m})$ et que tous les moments, s'ils existent, sont constants.

Pratiquement, comme on ne vérifiera jamais ces conditions, on définit une forme plus simple et plus facilement utilisable de la stationnarité : les processus covariances stationnaires. On dit qu'un processus stochastique discret est covariance stationnaire si sa moyenne et sa variance existent et sont constantes.

1.1.3. Processus stochastiques discrets linéaires

Tout processus stochastique discret pouvant être écrit sous la forme :

$$X_t = \mu + U_t + \psi_1 U_{t-1} + \psi_2 U_{t-2} + \dots$$

où μ et les ψ_i sont des paramètres et la suite $\{U_t; t \in \mathbb{Z}\}$ est composée de v.a. indépendantes et identiquement distribuées, de moyenne nulle et de variance σ_u^2 , est appelé processus stochastique discret linéaire.

Pour qu'un tel processus soit covariance stationnaire, il est suffisant que sa moyenne existe et soit constante et que sa matrice des covariances existe et soit indépendante de t .

Calculons ces éléments :

$$t, E(X_t) = \mu + E \sum_{i=0}^{\infty} \psi_i U_{t-i} \quad \text{avec } \psi_0 = 1$$

$$= \mu + E(U) \sum_{i=0}^{\infty} \psi_i$$

ce qui implique que la série $\sum_{i=0}^{\infty} \psi_i$ soit convergente
 auquel cas $E(X_t) = \mu$

$$\begin{aligned} V(X_t) &= E(X_t - E(X_t))^2 \\ &= E(U_t + \psi_1 U_{t-1} + \psi_2 U_{t-2} + \dots)^2 \\ &= E(U_t^2 + \psi_1^2 U_{t-1}^2 + \psi_2^2 U_{t-2}^2 + \dots) + E(\text{prod. croisés}) \\ &= E(U^2) \sum_{i=0}^{\infty} \psi_i^2 \quad \text{avec } \psi_0 = 1 \end{aligned}$$

ce qui implique que la série $\sum_{i=0}^{\infty} \psi_i^2$ doit être conver-

gente auquel cas $V(X_t) = \sigma_u^2 k$ ou $k = \lim_{n \rightarrow \infty} \sum_{i=0}^n \psi_i^2$

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))] \\ &= E[(U_t + \psi_1 U_{t-1} + \psi_2 U_{t-2} + \dots)(U_{t+h} + \psi_1 U_{t+h-1} + \psi_2 U_{t+h-2} + \dots)] \\ &= E\left[\sum_{i=0}^{\infty} \psi_i \psi_{i+h} U_{t-i}^2\right] \quad \text{avec } \psi_0 = 1 \\ &= E(U^2) \sum_{i=0}^{\infty} \psi_i \psi_{i+h} \end{aligned}$$

ce qui implique que la série $\sum_{i=0}^{\infty} \psi_i \psi_{i+h}$ doit être convergente pour tout n auquel cas $\text{Cov}(X_t, X_{t+h}) = \sigma_u^2 C(h)$ où $C(h) = \lim_{h \rightarrow \infty} \sum_{i=0}^h \psi_i \psi_{i+h}$ avec $\psi_0 = 1$

1.1.4. Processus autorégressifs

Tout processus stochastique s'écrit sous la forme :

$$X_t = \delta + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \dots + \theta_p X_{t-p} + U_t$$

où δ et les θ_i sont des paramètres et U_t est une v.a. de moyenne nulle et de variance σ_u^2 et telle que $\text{Cov}(U_t, U_{t+k}) = 0$ $k \in \mathbb{Z}$ est appelé processus autorégressif d'ordre p et noté AR(p).

I.1.5. Processus AR(1)

Un processus AR(1) s'écrit sous la forme :

$$\begin{aligned} X_t &= \delta + \theta X_{t-1} + U_t \\ &= \delta + \theta(\theta X_{t-2} + \theta U_{t-1}) + U_t \\ &\vdots \\ &= \theta^k X_{t-k} + \sum_{i=0}^k \theta^i U_{t-i} + \delta \sum_{i=0}^k \theta^i \end{aligned}$$

Les conditions de stationnarité imposent que $E(X_t)$ doit exister et être constant. Or $E(X_t) = \theta^k E(X_{t-k}) + E(U) \sum_{i=0}^k \theta^i + \delta \sum_{i=0}^k \theta^i$. Puisque k peut être choisi arbitrairement grand, il est nécessaire que la série

$\sum_{i=0}^{\infty} \theta^i$ converge, donc que $|\theta| < 1$.

Si tel est le cas X_t peut être écrit sous la forme d'un processus stochastique discret linéaire avec :

$$E(X_t) = \mu = \delta \sum_{i=0}^{\infty} \theta^i = \frac{\delta}{1-\theta}$$

$$V(X_t) = \sigma_u^2 \sum_{i=0}^{\infty} \theta^{2i} = \frac{\sigma_u^2}{1-\theta^2}$$

$$\text{Cov}(X_t, X_{t+h}) = \sigma_u^2 \theta^h \sum_{i=0}^{\infty} \theta^{2i} = \frac{\sigma_u^2 \theta^h}{1-\theta^2}$$

Comme on s'intéresse généralement au paramètre μ du processus et non au paramètre σ , on peut re-

paramétriser ce modèle et l'écrire sous la forme :

$$X_t - \mu = \theta(X_{t-1} - \mu) + U_t$$

Selon les besoins on utilisera une ou l'autre de ces 2 formes.

I.1.6. Génération d'un processus AR(1) stationnaire

Observer une réalisation d'un processus AR(1) revient à observer une suite de valeurs $\{x_i, i=1,2,\dots,n\}$ extraites arbitrairement d'une suite de valeurs $\{x_t, t \in \mathbb{Z}\}$. Donc lorsqu'on désire simuler une réalisation d'un processus stationnaire, il faut générer une valeur initiale x_1 qui soit tirée de la loi marginale commune à tous les X_t , car si la loi marginale de X_1 étoit différente de la loi marginale des autres X_t , le processus serait dans un état transitoire pendant un certain temps. On peut toutefois concevoir un processus AR(1) fini, commençant à $t=1$ mais en gardant à l'esprit qu'il doit être considéré comme la continuation d'un processus qui vient de $t=-\infty$.

Pour qu'un processus AR(1) généré à partir de $t=1$ soit covariance stationnaire, il faut donc que

$$E(X_1) = \mu = \frac{\delta}{1-\theta} \quad \text{et}$$

$$V(X_1) = \frac{\sigma_u^2}{1-\theta^2}$$

Il suffit donc que X_1 soit définie ainsi :

$$X_1 = \mu + \frac{U_1}{\sqrt{1-\theta^2}}$$

Avant de passer à l'estimation robuste des paramètres d'un processus AR(1), nous allons étudier le comportement de l'estimateur du maximum de vraisemblance (MLE) pour le cas de la loi de Gauss, ce qui nous amènera à l'estimateur des moindres carrés utilisé dans la pratique.

I.1.7. Estimation classique des paramètres du modèle AR(1)

Soit le modèle autarégressif

$$(X_t - \mu) = \theta(X_{t-1} - \mu) + U_t$$

où $|\theta| < 1$ et U_t sont des v.a. indépendantes de loi $g(x)$ quel que soit t . On se propose de calculer les MLE de μ , θ et σ^2 fondés sur un échantillon aléatoire X_1, X_2, \dots, X_n .

La loi de probabilité de $(X_t - \mu)$ conditionnée sur $(X_{t-1} - \mu)$ est la loi de probabilité de U_t puisque

$$f((x_t - \mu) | x_{t-1}) = f(u_t + \theta(x_{t-1} - \mu) | x_{t-1}) = g(u_t)$$

La loi de probabilité conjointe

$f((x_1-\mu), (x_2-\mu), \dots, (x_n-\mu))$ sera donc :

$$f(x_1-\mu) f((x_2-\mu) | x_1) f(x_3-\mu | x_1, x_2) \dots \\ \dots f(x_n-\mu | x_1, x_2, \dots, x_{n-1})$$

c'est-à-dire $f(x_1-\mu)g(u_2)g(u_3)\dots g(u_n)$

Dans le cas particulier de la loi de Gauss, c'est-à-dire si $U_i \sim \mathcal{N}(\mu_u=0 ; V_u = \sigma_u^2)$ alors

$$(X_1-\mu) \sim \mathcal{N}(\mu_1=0 ; V_1 = \frac{\sigma_u^2}{1-\theta^2}) \\ (X_2-\mu) | X_1 \sim \mathcal{N}(\mu_2=\theta(x_1-\mu) ; V_2 = \sigma_u^2) \\ \vdots \\ (X_n-\mu) | X_{n-1} \sim \mathcal{N}(\mu_n=\theta(x_{n-1}-\mu) ; V_n = \sigma_u^2)$$

Ainsi $f(x_1-\mu, x_2-\mu, \dots, x_n-\mu)$ vaut

$$\frac{(1-\theta^2)^{1/2}}{(2\pi)^{11/2} (\sigma^2)^{11/2}} \exp \left\{ -\frac{1}{2} (x_1-\mu)^2 \frac{1-\theta^2}{\sigma_u^2} + \right. \\ \left. + \frac{1}{\sigma_u^2} \sum_{i=2}^n [(x_i-\mu) - \theta(x_{i-1}-\mu)]^2 \right\}$$

Le lag de la fonction de vraisemblance vaut

alors

$$-\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_u^2 + \frac{1}{2} \ln (1-\theta^2) \\ - \frac{1}{2\sigma_u^2} \left\{ (x_1-\mu)^2 (1-\theta^2) + \sum_{i=2}^n [(x_i-\mu) - \theta(x_{i-1}-\mu)]^2 \right\}$$

Les solutions du maximum de vraisemblance sont difficiles à obtenir dans ce cas général à cause de la loi de $X_1 - \mu$. On se contentera donc de donner une solution à la fonction de vraisemblance basée sur la probabilité conditionnelle de $X_n - \mu$, $X_{n-1} - \mu, \dots, X_2 - \mu | X_1 = x_1$ qui, faisant disparaître les termes dûs à $f(x_1 - \mu)$ s'obtient facilement. On notera que pour n grand cette solution conditionnelle est très proche de la solution générale ou, plus précisément, elle lui est équivalente quand $n \rightarrow \infty$.

Le ln de la fonction de vraisemblance conditionnelle est donc :

$$\ln L = -\frac{(n-1)}{2} \ln(2\pi) - \frac{(n-1)}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \sum_{i=2}^n [(x_i - \mu) - \theta(x_{i-1} - \mu)]^2$$

d'où

$$\frac{\partial}{\partial \mu} \ln L = \frac{1}{\sigma_u^2} \sum_{i=2}^n [(x_i - \mu) - \theta(x_{i-1} - \mu)](1 - \theta)$$

$$\frac{\partial}{\partial \theta} \ln L = \frac{1}{\sigma_u^2} \sum_{i=2}^n [(x_i - \mu) - \theta(x_{i-1} - \mu)](x_{i-1} - \mu)$$

De la première dérivée on en déduit

$$\hat{\mu} = \frac{\sum_{i=2}^n X_i - \hat{\theta} \sum_{i=2}^n X_{i-1}}{1 - \hat{\theta}}$$

et de la seconde

$$\hat{\theta} = \frac{\sum_{i=2}^n (X_i - \hat{\mu})(X_{i-1} - \hat{\mu})}{\sum_{i=2}^n (X_{i-1} - \hat{\mu})^2}$$

c'est-à-dire, après avoir substitué $\hat{\mu}$ par sa valeur et en écrivant

$$\bar{X}_{-i} = \frac{\sum_{j=1}^n X_j - X_i}{n}$$

$$\hat{\theta} = \frac{\sum_{i=2}^n (X_i - \bar{X}_{-1})(X_{i-1} - \bar{X}_{-n})}{\sum_{i=2}^n (X_{i-1} - X_{-n})^2}$$

Ces estimateurs sont les estimateurs classiques des moindres carrés puisque l'on a minimisé

$$\sum_{i=2}^n [(X_i - \mu) - \theta(X_{i-1} - \mu)]^2 .$$

Remarques :

- à l'ordre $\frac{1}{n}$ près $\hat{\mu}$ est \bar{X} et $\hat{\theta}$ est le premier terme du corrélogramme;
- on obtient également $\hat{\mu}$ et $\hat{\theta}$ en résolvant l'équation de régression classique $X_t = \delta + \theta X_{t-1} + U_t$ et en estimant $\hat{\mu}$ par $\frac{\hat{\delta}}{1-\hat{\theta}}$.

Pour obtenir le MLE de σ_u^2 , on dérive $\ln L$ par rapport à σ_u^2 , soit :

$$\frac{\partial}{\partial \sigma_u^2} \ln L = -\frac{n-1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_{i=2}^n [(x_i - \mu) - \theta(x_{i-1} - \mu)]^2$$

et on obtient que

$$\hat{\sigma}_u^2 = \frac{1}{n-1} \sum_{i=2}^n [(X_i - \hat{\mu}) - \hat{\theta}(X_{i-1} - \hat{\mu})]^2$$

I.1.8. Variances asymptotiques des estimateurs des moindres carrés

En posant $\hat{\Delta}_n^T = (\hat{\delta} \hat{\theta})$ et

$$X_n^T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_2 & X_3 & X_4 & \dots & X_n \end{bmatrix} \text{ et en tenant compte}$$

du fait que (Fuller, 1976, p. 335) :

$$\lim_{n \rightarrow \infty} n V(\hat{\Delta}_n) = \sigma_u^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n-1} X_n^T X_n \right)^{-1}$$

on trouve que

$$\begin{aligned} \lim_{n \rightarrow \infty} n V(\hat{\Delta}_n) &= \sigma_u^2 \begin{bmatrix} 1 & \frac{\delta}{1-\theta} \\ \frac{\delta}{1-\theta} & \frac{\sigma_u^2}{1-\theta^2} + \frac{\delta^2}{(1-\theta)^2} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \sigma_u^2 + \delta^2 \frac{(1+\theta)}{1-\theta} & -\delta(1+\theta) \\ -\delta(1+\theta) & 1-\theta^2 \end{bmatrix} \end{aligned}$$

Ainsi la variance asymptotique de $n \hat{\delta}$ est égale à $\sigma_u^2 + \delta^2 \frac{(1+\theta)}{1-\theta} = \sigma_u^2 + \mu^2 (1-\theta^2)$, la variance asymptotique de $n \hat{\theta}$ vaut $1-\theta^2$ et asymptotiquement $n \text{cov}(\hat{\delta}, \hat{\theta}) = -\frac{\delta(1-\theta^2)}{1-\theta}$.

Asymptotiquement, comme $\hat{\mu}$ est équivalent à \bar{X} , $n V(\hat{\mu}) = n V(\bar{X}) = \sum V(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) =$

$$= \frac{\sigma_u^2}{(1-\theta)^2}$$

Du point de vue de la robustesse ces résultats sont intéressants. On remarque d'abord que la variance asymptotique de $\hat{\theta}$ ne dépend pas de U ou d'une fonction de U . Cela signifie que quelle que soit la distribution de l'erreur U_t du modèle autorégressif et donc quelle que soit la variance σ_u^2 de U_t la méthode des moindres carrés fournira un estimateur de θ dont la variance asymptotique ne dépend que de θ . Donc $\hat{\theta}$ est qualitativement (au sens de Hampel) très robuste. Par contre autant la variance asymptotique de $\hat{\delta}$ que de $\hat{\mu}$ dépendent de σ_u^2 . Autrement dit ces estimateurs ne sont pas robustes au sens de Hampel puisque leur variance asymptotique est proportionnelle à σ_u^2 .

Un autre aspect de la robustesse nous intéresse : la robustesse au sens de l'efficacité qui est la notion de robustesse au sens de Huber. Lorsque la loi des erreurs est gaussienne, on sait que les estimateurs des moindres carrés de θ et μ sont asymptotiquement efficaces (efficacité=1) puisqu'ils sont MLE. Mais si la loi des erreurs n'est pas gaussienne, que devient l'efficacité de ces estimateurs ? Autrement dit, on veut savoir comment l'efficacité de ces estimateurs dépend de la loi des erreurs U_t .

I.1.9. Efficacité des estimateurs des moindres carrés

Pour trouver l'efficacité des estimateurs des moindres carrés de μ et θ , comparons leur variance asymptotique à la variance asymptotique minimale atteignable pour des estimateurs convergents, c'est-à-dire à la borne inférieure de Cramer-Roo pour une loi des U_t quelconque.

Considérons d'abord l'efficacité de $\hat{\theta}$ en posant le modèle stationnaire supposé centré ($\mu=0$) sans restreindre la généralité :

$$X_t = \theta X_{t-1} + U_t$$

avec $|\theta| < 1$ et U_t une suite de v.a. indépendantes et identiquement distribuées de loi g .

Pour un échantillon aléatoire X_1, X_2, \dots, X_n la densité de probabilité conjointe vaut :

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_n | x_{n-1}, \dots, x_1; \theta) \cdot f(x_{n-1} | x_{n-2}, \dots, x_1, \theta) \dots f(x_1; \theta)$$

En conséquence le logarithme de la fonction de vraisemblance de θ est égal à :

$$l(\theta; x_1, x_2, \dots, x_n) = \sum_{j=2}^n \ln f(x_j | x_{j-1}, \dots, x_1; \theta) + \ln f(x_1; \theta)$$

Soit $I(\theta)$ l'information de Fisher fondée sur

l'échantillon X_1, X_2, \dots, X_n . On a que

$$I(\theta) = \sum_{j=2}^n E\left[-\frac{\partial^2}{\partial \theta^2} \ln f(x_j | x_{j-1}, \dots, x_1; \theta)\right] + E\left[-\frac{\partial^2}{\partial \theta^2} \ln f(x_1; \theta)\right]$$

soit en posant $I_j(\theta)$ pour l'information due à X_j conditionnée sur X_{j-1}, \dots, X_1

$$I(\theta) = \sum_{j=2}^n I_j(\theta) + I_1(\theta) \text{ ou } I_1(\theta) = E\left[-\frac{\partial^2}{\partial \theta^2} \ln f(x_1; \theta)\right]$$

Calculons $I_j(\theta)$ pour $j \geq 2$. Puisque $f(x_j | x_{j-1}, \dots, x_1; \theta)$ est la densité conditionnelle de X_j , elle vaut $g(u_j) = g(x_j - \theta x_{j-1})$. Ainsi

$$\begin{aligned} I_j(\theta) &= E\left[-\frac{\partial^2}{\partial \theta^2} \ln g(x_j - \theta x_{j-1})\right] \\ &= \iint_{-\infty}^{+\infty} \frac{x_{j-1}^2 g'^2(x_j - \theta x_{j-1}) - x_{j-1}^2 g''(x_j - \theta x_{j-1}) g(x_j - \theta x_{j-1})}{g^2(x_j - \theta x_{j-1})} \\ &\quad \cdot f(x_j, x_{j-1}) dx_j dx_{j-1} \end{aligned}$$

$$\begin{aligned} \text{Or } f(x_j, x_{j-1}) &= f(x_j | x_{j-1}, \dots, x_1; \theta) f(x_{j-1}; \theta) \\ &= g(u_j) f(x_{j-1}; \theta) \end{aligned}$$

Donc

$$\begin{aligned} I_j(\theta) &= \iint_{-\infty}^{+\infty} \frac{x_{j-1}^2 g'^2(u_j) - x_{j-1}^2 g''(u_j) g(u_j)}{g^2(u_j)} \\ &\quad \cdot g(u_j) f(x_{j-1}; \theta) du_j dx_{j-1} \end{aligned}$$

$$= \int_{-\infty}^{+\infty} x_{j-1}^2 f(x_{j-1}; \theta) dx_{j-1} \int_{-\infty}^{+\infty} \frac{g'^2(u_j) - g''(u_j)g(u_j)}{g^2(u_j)} g(u_j) du_j$$

$$= E[X_{j-1}^2] E\left[-\frac{d^2}{du_j^2} \ln g(u_j)\right] = \frac{\sigma_u^2}{1-\theta^2} E\left[-\frac{d^2}{du_j^2} \ln g(u_j)\right]$$

que nous noterons $\frac{\sigma_u^2}{1-\theta^2} I_g$

Ainsi

$$I(\theta) = (n-1) I_g \frac{\sigma_u^2}{1-\theta^2} + I_1 = n I_g \frac{\sigma_u^2}{1-\theta^2} + o(1)$$

Donc l'efficacité asymptotique de l'estimateur des moindres carrés vaudra

$$\frac{1}{\sigma_u^2 I_g}$$

On retrouve que dans le cas d'une distribution gaussienne des erreurs U_t , l'efficacité est égale à 1 puisque dans ce cas $I_g = \frac{1}{\sigma_u^2}$.

Comme le fait remarquer Martin (1976, 1978), cette efficacité n'est pas robuste car σ_u^2 peut augmenter considérablement si la loi de U est allongée dans ses queues alors que I_g peut rester beaucoup plus stable.

Pour le paramètre de translation μ du modèle $X_{t-\mu} = \theta(X_{t-1-\mu}) + U_t$, on trouve par un calcul analo-

gue que $I(\mu) = n(1-\theta)^2 I_g$. Donc l'efficacité asymptotique de l'estimateur des moindres carrés de μ est égale à

$$\frac{1}{n(1-\theta)^2 I_g} / \frac{\sigma_u^2}{(1-\theta)^2} = \frac{1}{\sigma_u^2 \cdot I_g}$$

c'est-à-dire qu'elle est identique à celle de l'estimateur des moindres carrés de θ .

Nous sommes donc amenés à rechercher des estimateurs qui soient robustes au sens de l'efficacité, c'est-à-dire des estimateurs capables de diminuer la variance de U_t lorsque l'on s'écarte du modèle de distribution gaussien des erreurs. Nous y parviendrons en appliquant des M-estimateurs à U_t , c'est-à-dire à $(X_t - \mu) - \theta(X_{t-1} - \mu)$ puisque nous savons que ces estimateurs sont capables de réduire la variance asymptotique de U_t lorsque la loi de U_t est contaminée.

Auparavant, nous allons décrire deux modèles de contamination généralement utilisés dans l'étude des processus autorégressifs contaminés et choisir un modèle pour la suite de cette étude.

1.2. Types de modèles AR(1) contaminés

Martin et Jong (1976), Denby et Martin (1979) ont défini et étudié deux modèles autorégressifs contaminés simples mais capables de représenter certaines situations réelles avec précision.

a) Le modèle de perturbations innovatrices

Ce modèle s'écrit :

$$X_t - \mu_v = \theta(X_{t-1} - \mu_v) + U_t \quad (1.2.1)$$

$$Y_t = X_t \quad (1.2.2)$$

Dans ce modèle U_t est une suite de v.o. indépendantes et identiquement distribuées par une loi de Gauss contaminée. Le processus qui se déroule est décrit par l'équation (1.2.1) et le processus qu'on observe est Y_t défini par l'équation (1.2.2). Lorsque la proportion de la contamination est faible, l'apparition d'une erreur importante aura une influence durable sur la moyenne du processus. Bien des séries temporelles en économie peuvent être représentées par un tel modèle. Par exemple une série de production journalière ou hebdomadaire perturbée tout-à-coup par des grèves; ou bien une série d'un indice boursier perturbé par une faillite, une décision de politique économique de la part d'un gouvernement ou encore par l'éclatement de conflits dans certains pays.

En fait, chaque fois que l'on observe de manière exacte, c'est-à-dire, sans erreurs systématiques un processus AR(1), mais que l'on se trompe ou que l'on risque de se tromper dans la spécification des erreurs (par exemple en affirmant que la distribution des erreurs est gaussienne alors qu'en réalité elle ne l'est pas) un tel modèle convient bien pour décrire la réalité.

b) Le modèle de perturbations interruptives

Ce modèle s'écrit :

$$X_t - \mu_v = \theta(X_{t-1} - \mu_v) + U_t \quad (1.2.3)$$

$$Y_t = X_t + V_t \quad (1.2.4)$$

Dans ce modèle U_t est une suite de v.a. indépendantes identiquement distribuées par une loi non contaminée. V_t quant à elle est une suite de v.a. indépendantes identiquement distribuées par une loi de la forme $(1-\gamma)\delta_0(x) + \gamma H(x)$

$$\text{où } \delta_0(x) = \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{si } x \neq 0 \end{cases}, \quad H(x) \text{ est une}$$

densité contaminatrice et γ est la proportion de contamination généralement inférieure à 0.10. Afin de diminuer les difficultés théoriques, on ajoute généralement l'hypothèse que U_t et V_t sont des suites de v.a. indépendantes entre elles. Le processus décrit par

l'équation (I.2.3) se déroule donc de manière exacte, mais son observation Y_t définie par l'équation (I.2.4) est entachée d'erreurs qui se produisent de temps à autres sans perturber de manière durable le déroulement du processus. Ici également, bien des séries réelles peuvent être représentées par un tel modèle, notamment lorsque des données ont été mal enregistrées ou lorsqu'un phénomène extérieur au processus est venu passagèrement perturber son déroulement.

Nous étudierons le comportement des M-estimateurs pour le modèle des perturbations innovatrices lorsque la contamination est asymétrique. Le comportement des M-estimateurs pour le modèle des perturbations interruptives a été étudié par Martin et Jong (1976) pour des contaminations symétriques. Leurs résultats montrent que les M-estimateurs sont inadaptés à ce genre de modèle et qu'il faut alors utiliser des estimateurs d'une classe plus large : la classe des M-estimateurs généralisés. Comme cette étude porte sur les M-estimateurs, nous n'étudierons que le modèle des perturbations innovatrices.

1.3. Etude théorique en situation de contamination asymétrique

Comme nous l'avons fait en première partie pour estimer le paramètre de translation d'une loi contaminée, nous allons estimer les paramètres μ_v et θ du modèle AR(1) défini en I.2.a, pour un échantillon aléatoire X_1, X_2, \dots, X_n par les M-estimateurs M_n et T_n qui vérifient

$$\sum_{i=2}^n \rho \left[\frac{(X_i - M_n) - T_n (X_{i-1} - M_n)}{S_n} \right] = \min$$

où M_n est l'estimateur de μ_v et T_n l'estimateur de θ . Toutefois par commodité nous posons $D_n = M_n (1 - T_n)$ et nous estimerons, par analogie avec la régression robuste, les paramètres μ_v et θ par les estimateurs D_n et T_n vérifiant

$$\sum_{i=2}^n \rho \left(\frac{X_i - T_n X_{i-1} - D_n}{S_n} \right) = \min$$

Ainsi D_n et T_n seront solutions des équations

$$\sum_{i=2}^n \psi \left(\frac{X_i - T_n X_{i-1} - D_n}{S_n} \right) = 0$$

$$\sum_{i=2}^n X_{i-1} \psi \left(\frac{X_i - T_n X_{i-1} - D_n}{S_n} \right) = 0$$

S_n est l'estimateur de σ_u choisi de manière à standardiser la fonction ψ . Puisque $\mu_v = \frac{\delta_v}{1-\theta}$, $\frac{\delta}{1-\theta}$ estimera μ_v .

Comme auparavant nous nous limiterons au cas où $E(U_t)$ est positive ou nulle pour tout t , sans restreindre la généralité, car si $E(U_t)$ est négative, il suffira de choisir une nouvelle fonction $\psi_1(x)$ égale à $-\psi(-x)$.

On remarque que cette spécification de la moyenne des erreurs représente une situation analogue à celle qui fut étudiée en première partie où les estimateurs du paramètre de translation étaient généralement biaisés. Il en sera de même ici, lorsqu'on estimera le paramètre μ_v , moyenne du processus non contaminé.

Supposons que μ_v et θ sont les paramètres à estimer d'un modèle AR(1) et supposons d'autre part que $E(U_t) = \mu_a$ où U_t est l'erreur aléatoire au temps t . Prenons comme estimateurs de μ_v et θ les estimateurs des moindres carrés $\hat{\mu}_n$ et $\hat{\theta}_n$. Comme

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \frac{\text{Cov}(X_t, X_{t-1})}{V(X_{t-1})} = \theta, \quad \hat{\theta}_n \text{ sera asymptotique-}$$

ment sans biais. Par contre

$$\lim_{n \rightarrow \infty} \hat{\mu}_n = E(X_t) = \mu_v + \frac{\mu_a}{1-\theta}. \quad \text{Donc } \hat{\mu}_n \text{ admettra un}$$

biais asymptotique de valeur $\frac{\mu_0}{1-\theta}$. De manière plus générale si $\hat{\mu}_n$ et $\hat{\theta}_n$ sont des M-estimateurs de μ_v et θ , l'estimateur $\hat{\mu}_n$ sera généralement biaisé, même asymptotiquement. Par contre la propriété de convergence de $\hat{\theta}_n$ sera conservée pour $\hat{\theta}_n$, dû au fait, intuitivement, que $\psi(U_t)$ et $\psi(U_{t-j})$ restent indépendantes pour $j > 0$, et ceci quelle que soit la loi des erreurs. Toutefois $\hat{\theta}_n$ comme $\hat{\mu}_n$ sont biaisés pour des échantillons de taille finie. Le biais de $\hat{\theta}_n$ vaut en première approximation (Kendall et Stuart, 1976, vol.3, p.453 et s.) :

$$-\frac{1+3\theta}{n}$$

En ce qui concerne la variance asymptotique de ces estimateurs, nous nous référons à Martin (1981). Avec des hypothèses qui ne sont pas formulées de manière précise, Martin a calculé que la matrice des variances-covariances asymptotiques de

$$\begin{array}{l}
 \hat{\lambda}_n = \begin{bmatrix} \hat{\mu}_n \\ \hat{\theta}_n \end{bmatrix} \text{ vaut :} \\
 V_{\hat{\lambda}} = \begin{bmatrix} 1 + \frac{\mu_c^2(1-\theta^2)}{\sigma_u^2} & -\frac{\mu_c(1-\theta^2)}{\sigma_u^2} \\ -\frac{\mu_c(1-\theta^2)}{\sigma_u^2} & \frac{1-\theta^2}{\sigma_u^2} \end{bmatrix} V_{\text{LOC}}(\psi, g(u))
 \end{array}$$

où $V_{LOC} = \frac{E_g[\psi^2(U-C)]}{[\frac{\partial}{\partial C}\lambda'(C)]^2}$ est la formule déjà rencontrée

dans la première partie, au point I.1.3.

Comme nous nous intéressons habituellement au paramètre μ_v représentant la moyenne du processus non contaminé, nous donnons l'expression de la matrice V_M^v des variances-covariances asymptotiques des M-estimateurs $\hat{M}_n^v = (\hat{\mu}_n \hat{\theta}_n)$, où $\hat{\mu}_n = \frac{\hat{\gamma}_n}{1-\hat{\theta}_n}$

La matrice des Jacobiens représentant la transformation $M \rightarrow \Delta$ vaut :

$$H = \begin{bmatrix} 1 - \theta & -\mu \\ 0 & 1 \end{bmatrix}$$

La matrice V_M^v vaut alors (Martin, 1981)

$$V_M^v = H^{-1} V_{\Delta}^v H^T^{-1} = \begin{bmatrix} \frac{1}{(1-\theta)^2} & 0 \\ 0 & \frac{1-\theta^2}{\sigma_u^2} \end{bmatrix} V_{LOC}(\psi, g(u))$$

On vérifie qu'en prenant la fonction $\psi(x)=x$, $V_{LOC}=\sigma_u^2$, et l'on obtient alors la matrice des variances-covariances asymptotiques des estimateurs des moindres carrés $\hat{\mu}_n$ et $\hat{\theta}_n$. On remarque, fait intéressant, qu'asymptotiquement $\hat{\mu}_n$ et $\hat{\theta}_n$ sont non corrélés. On s'aperçoit également que la variance asymptotique

de $\hat{\theta}_n$ dépend de σ_u^2 ce qui n'était pas le cas pour $\hat{\theta}_n$. On perd donc en partie la propriété de robustesse qualitative attachée à $\hat{\theta}_n$. Ce fait n'est toutefois pas gênant, car on pourra ainsi obtenir des estimations de θ plus précises en utilisant des fonctions ψ qui stabilisent V_{LOC} alors que σ_u^2 augmente. Par contre, la variance asymptotique de $\tilde{\mu}_n$ dépend de V_{LOC} et non pas de σ_u^2 alors que celle de $\hat{\mu}_n$ était une fonction de σ_u^2 . Ainsi, même si σ_u^2 augmente considérablement, on arrivera à contenir la croissance de la variance de $\tilde{\mu}_n$. Evidemment, lorsque la contamination sera faible ou nulle V_{LOC} sera alors supérieure à σ_u^2 et l'on perdra une certaine précision dans l'estimation de θ comme dans celle de μ_v . C'est en quelque sorte le prix payé pour l'assurance que l'on prend afin de se prémunir contre des erreurs dont la variance peut être élevée.

Nous allons maintenant procéder à des simulations sur des échantillons de taille finie. Cela nous permettra de vérifier donc quelle mesure les résultats asymptotiques s'appliquent et de comparer les performances des estimateurs dans diverses situations de contamination.

Chapitre II. SIMULATIONS POUR DES ECHANTILLONS DE TAILLE FINIE

Les estimateurs comparés sont ceux qui furent proposés en conclusion de la première partie, à savoir :

$$S(1;0;1)k=1.4$$

$$S(0.75;0.25;1)k=1.25 \text{ et}$$

$$A(0.5;0;1)k=1.5$$

A ces estimateurs, l'estimateur des moindres carrés a été adjoint afin de comparer ses performances à celles des autres estimateurs.

Les contaminations prises en compte sont toutes des combinaisons convexes de deux distributions de Gauss de la forme $(1-\epsilon)\mathcal{N}(0,1)+\epsilon\mathcal{N}(\mu_0,\sigma_0^2)$, soit

<u>ϵ</u>	<u>μ_0</u>	<u>σ_0^2</u>	<u>nom symbolique</u>
0.0	0.0	1.0	G000
0.1	0.0	16.0	G010
0.1	2.0	1.0	G210
0.1	4.0	1.0	G410
0.2	4.0	1.0	G420

La contamination G420 a été ajoutée par rapport à la première partie afin de comparer les estimateurs dans des situations de contamination extrêmes

faisant apparaître distinctement deux modes.

Le modèle étudié est :

$$(X_t - 10) = 0.5(X_{t-1} - 10) + U_t$$

Nous n'avons pas effectué de simulations pour d'autres valeurs de ces paramètres car les résultats comparatifs restent valables quelles que soient ces valeurs.

A titre d'exemple, nous avons dessiné sur les figures II.1.1 à II.1.5 la réalisation d'un échantillon de taille 50 pour chacune des contaminations étudiées.

La génération des échantillons pour la simulation a été établie de la façon suivante :

- 1) générer un nombre au hasard entre 0 et 1;
- 2) si ce nombre est supérieur à ϵ , alors générer la réalisation d'une variable aléatoire de loi $\mathcal{N}(0,1)$, sinon générer une réalisation d'une variable aléatoire d'une loi $\mathcal{N}(\mu_0, \sigma_0^2)$. La valeur générée est notée z_0 ;
- 3) générer la première observation du processus

$$x_0 = \mu + \frac{\epsilon \mu_0}{1-\theta} + \frac{z_0 - \epsilon \mu_0}{\sqrt{1-\theta^2}}$$

On obtient ainsi une réalisation d'une v.a. qui a

même moyenne et même variance que les autres v.a.,
comme cela a été dit en I.1.6;

- 4) générer un nouveau nombre au hasard entre 0 et 1;
- 5) procéder de manière identique au point 2. La valeur générée est alors notée z_i ;
- 6) générer l'observation

$$x_i = 0.5 x_{i-1} + 5.0 + \epsilon_i$$

- 7) recommencer au point 4) n fois plus 20 ou n représente la longueur désirée de l'échantillon;
- 8) on ne conserve que les n dernières observations.
Ainsi, en plus des conditions imposées sur x_0 afin que la série obtenue soit stationnaire, on prend une précaution supplémentaire en éliminant les 20 premières observations.

Le générateur de nombres au hasard est celui de "Mathematical Software ACM vol. 5, No 2, p. 132 - 138". Les observations d'une loi de Gauss ont été obtenues à l'aide de fonctions congruentes.

L'estimation des paramètres a été obtenue selon les équations données en I.3 pour D_n et T_n , par un sous-programme inclus dans la librairie ROBETH. Cette librairie contient des sous-programmes écrits par A. Marazzi, ETH, ZÜRICH, permettant d'effectuer diverses procédures statistiques robustes.

Les longueurs d'échantillons retenues ont été

20, 50 et 100. Chaque situation d'échantillonnage a été générée 600 fois. On a alors calculé la moyenne des estimations de μ_V et θ ainsi que les variances empiriques.

Afin de contrôler la stabilité des résultats, la simulation a été opérée deux fois. Nous avons calculé les variations en % des valeurs obtenues au cours des deux simulations, sous la forme $\frac{R_{MAX} - R_{MIN}}{R_{MAX}} \cdot 100$

où R_{MAX} et R_{MIN} sont respectivement la valeur la plus grande et la valeur la plus petite des 2 simulations. Le tableau II.1.1 représente les variations maximales pour chaque taille d'échantillon en mettant en évidence pour quel estimateur et quelle contamination cette variation maximale est apparue.

Finalement, nous avons agrégé les 2 fois 600 échantillons en les considérant comme 1200 échantillons et en calculant la moyenne et la variance comme pour une simulation faite avec 1200 échantillons. Les résultats sont présentés dans les tableaux II.1.2 à II.1.5. Les variances ont été multipliées par la longueur de l'échantillon. Toutes les valeurs ont été arrondies à la deuxième décimale.

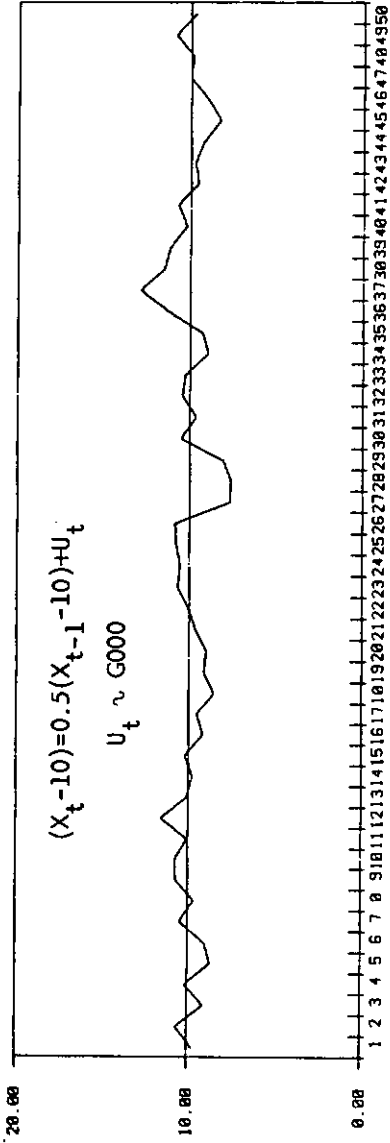


Figure II.1.1

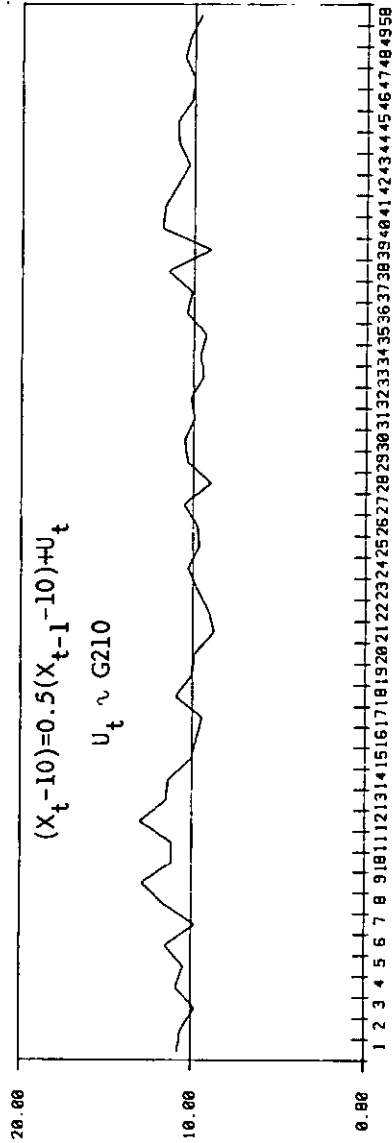


Figure II.1.2

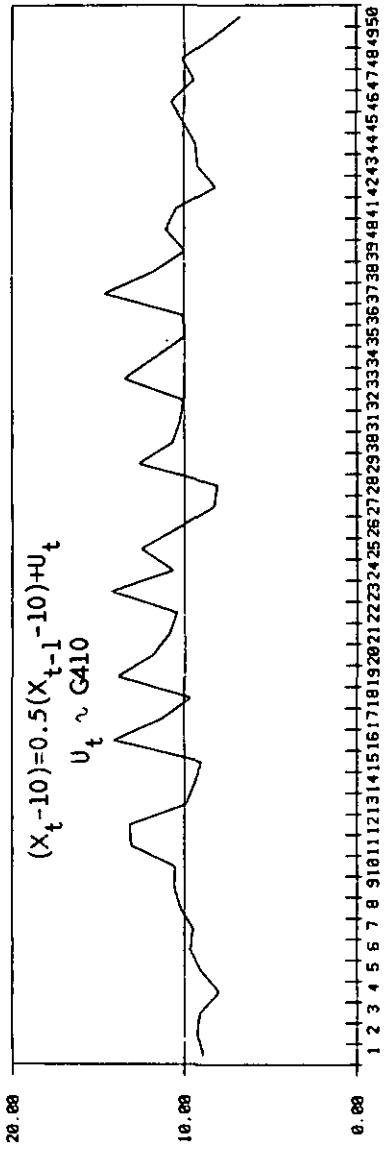


Figure II.1.3

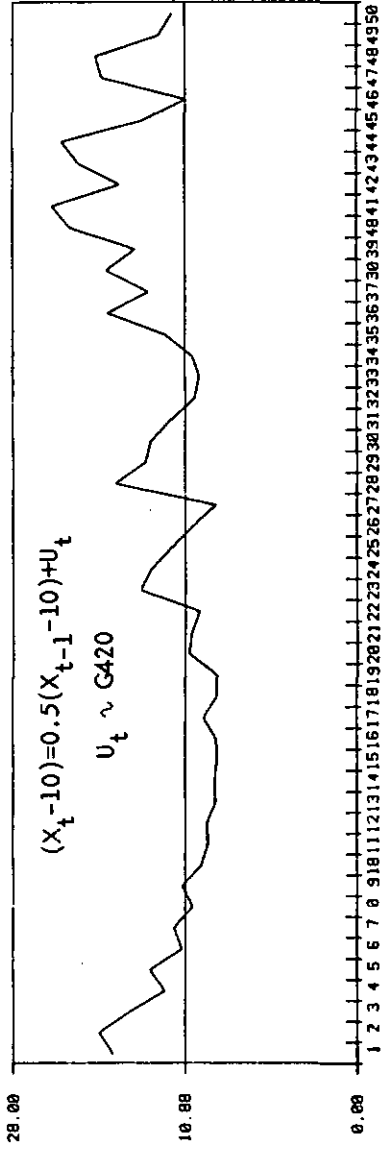


Figure II.1.4

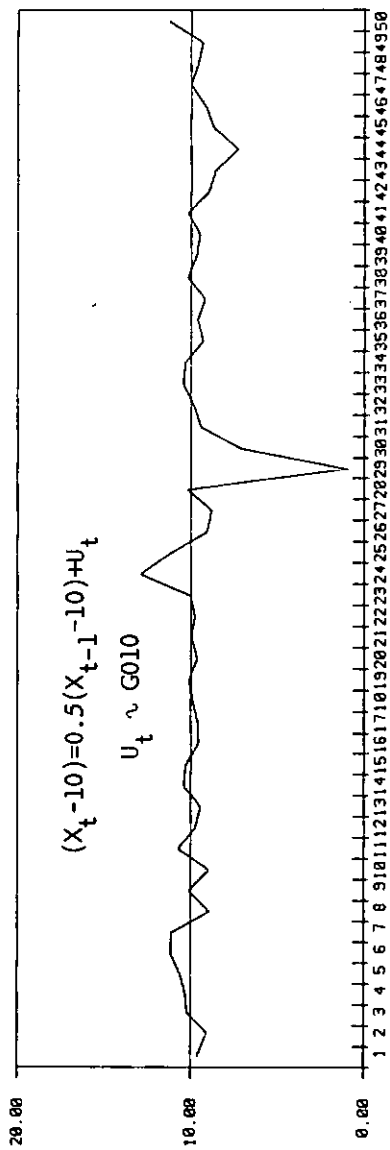


Figure 11.1.5

variations maximales						
	20	estimeur contamination	50	estimeur contamination	100	estimeur contamination
μ_n	MOY	0.64 A(0.5;0;1)k=1.5 G420	0.89	A(0.5;0;1)k=1.5 G010	0.22	A(0.5;0;1)k=1.5 G010
	VAR	33.89 S(0.75;0.25;1)k=1.25 G210	8.62	A(0.5;0;1)k=1.5 G010	8.45	A(0.5;0;1)k=1.5 G010
θ_n	MOY	2.63 S(1;0;1)k=1.4 G420	2.12	S(0.75;0.25;1)k=1.25 G000	1.25	S(0.75;0.25;1)k=1.25 G000
	VAR	8.64 S(0.75;0.25;1)k=1.25 G420	8.60	G(0.75;0.25;1)k=1.25 G000	8.33	S(0.75;0.25;1)k=1.25 G210

Tableau II.1.1.1

Estimation des paramètres μ_y et θ du modèle $X_t - 10 = 0.5(X_{t-1} - 10) + U_t$
 avec l'estimateur $S(1;0;1)_{k=\infty}$ (les variances ont été multipliées par n)

taille échant.		10		20		50		∞	
		$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	μ
G000	MOY.	0.37	10.02	0.45	10.00	0.48	10.00	0.50	10.00
	VAR.	0.87	4.11	0.82	3.97	0.77	4.14	0.75	4.00
G010	MOY.	0.38	10.00	0.45	10.00	0.48	10.00	0.50	10.00
	VAR.	0.77	12.49	0.73	9.70	0.75	9.79	0.75	10.00
G210	MOY.	0.38	10.42	0.46	10.40	0.48	10.40	0.50	10.40
	VAR.	0.84	5.56	0.76	5.34	0.79	5.37	0.75	5.44
G410	MOY.	0.38	10.80	0.45	10.80	0.47	10.80	0.50	10.80
	VAR.	0.80	17.47	0.72	10.52	0.75	9.42	0.75	9.76
G420	MOY.	0.38	11.61	0.45	11.60	0.48	11.60	0.50	11.60
	VAR.	0.86	14.81	0.79	13.80	0.78	14.29	0.75	14.24

Tableau II.1.2

Estimation des paramètres μ_y et θ du modèle $X_t - 10 = 0.5(X_{t-1} - 10) + U_t$
 avec l'estimateur $S(1;0;1)k=1.4$ (les variances ont été multipliées par n)

taille échant.		20		50		100		∞	
		θ	$\hat{\mu}$	θ	$\hat{\mu}$	θ	$\hat{\mu}$	θ	$\hat{\mu}$
G000	MOY.	0.38	10.00	0.45	9.99	0.48	10.00	0.50	10.00
	VAR.	0.86	4.60	0.84	4.08	0.80	4.26	0.78	4.18
G010	MOY.	0.40	10.02	0.46	9.99	0.48	10.00	0.50	10.00
	VAR.	0.67	7.91	0.55	6.19	0.48	5.78	0.43	5.73
G210	MOY.	0.38	10.35	0.45	10.34	0.48	10.33	0.50	10.33
	VAR.	0.89	5.39	0.81	5.46	0.75	5.36	0.74	5.37
G410	MOY.	0.40	10.60	0.46	10.55	0.48	10.51	0.50	10.50
	VAR.	0.68	8.98	0.58	8.45	0.58	8.53	0.55	7.14
G420	MOY.	0.39	11.36	0.46	11.32	0.48	11.30	0.50	11.27
	VAR.	0.78	16.47	0.73	15.83	0.73	16.76	0.72	13.74

Tableau II.1.3

Estimation des paramètres μ_v et θ du modèle $X_t - 10 = 0.5(X_{t-1} - 10) + U_t$
 avec l'estimateur $S(0.75; 0.25; 1)k=1.25$ (les variances ont été multipliées par n)

taille échant.		20		50		100		∞	
		θ	$\hat{\mu}$	θ	$\hat{\mu}$	θ	$\hat{\mu}$	θ	$\hat{\mu}$
G000	MOY.	0.38	10.01	0.45	10.00	0.48	10.00	0.50	10.00
	VAR.	0.94	4.71	0.93	4.40	0.86	4.56	0.85	4.50
G010	MOY.	0.41	9.99	0.46	10.00	0.48	10.00	0.50	10.00
	VAR.	0.71	6.59	0.57	6.00	0.47	5.97	0.40	5.32
G210	MOY.	0.39	10.34	0.45	10.31	0.47	10.31	0.50	10.30
	VAR.	0.88	15.29	0.81	5.60	0.84	5.49	0.78	5.63
G410	MOY.	0.40	10.49	0.46	10.43	0.48	10.41	0.50	10.40
	VAR.	0.65	10.02	0.57	7.60	0.52	7.73	0.50	6.57
G420	MOY.	0.40	11.13	0.46	11.04	0.48	11.00	0.50	10.98
	VAR.	0.74	16.54	0.65	14.76	0.61	13.94	0.60	11.48

Tableau II.1.4

Estimation des paramètres μ_y et θ du modèle $X_t - 10 = 0.5(X_{t-1} - 10) + U_t$
 avec l'estimateur $A(0.5; 0; 1)_{k=1.5}$ (les variances ont été multipliées par n)

toille échant.		20		50		100		∞	
		$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{\mu}$	$\hat{\theta}$	$\hat{\mu}$
G000	MOY.	0.38	9.45	0.45	9.41	0.48	9.37	0.50	9.36
	VAR.	0.97	5.92	0.93	5.65	0.86	5.17	0.83	4.44
G010	MOY.	0.40	9.22	0.46	9.20	0.48	9.15	0.50	9.14
	VAR.	0.75	25.62	0.84	14.39	0.78	10.77	0.82	10.95
G210	MOY.	0.38	9.76	0.46	9.69	0.48	9.67	0.50	9.65
	VAR.	0.84	6.91	0.81	6.93	0.73	6.12	0.72	5.20
G410	MOY.	0.41	9.90	0.46	9.87	0.48	9.83	0.50	9.81
	VAR.	0.62	9.08	0.48	8.01	0.45	7.81	0.46	6.04
G420	MOY.	0.42	10.50	0.46	10.47	0.48	10.41	0.50	10.37
	VAR.	0.59	13.54	0.54	13.31	0.56	13.13	0.50	9.48

Tableau II.1.5

Commentaires généraux

Pour les 4 estimateurs considérés, la moyenne de $\hat{\mu}_n$ est proche de sa valeur asymptotique, déjà dans les échantillons de taille 20, pour toutes les situations de contamination. La différence la plus grande est obtenue avec l'estimateur $S(0.75;0.25;1)k=1.25$ pour la situation G420 où la moyenne de $\hat{\mu}_{20}$ vaut 11.13, ce qui représente une différence de moins de 2 % par rapport à la valeur asymptotique. De plus les estimations obtenues par les moindres carrés sont toujours les plus proches de leur valeur asymptotique. Pour les échantillons de taille 20 déjà, cette différence est remarquablement faible.

Quant à la variance $\hat{\mu}_n$, on s'aperçoit que les variations sont plus importantes. Les variances pour les échantillons de taille 20 ne sont pas supérieures de plus de 20 % aux valeurs asymptotiques dans les cas suivants : G000, G210, G420 pour l'estimation par les moindres carrés; G000 pour l'estimation avec $S(0.75;0.25;1)k=1.25$; G000, G210, G410 et G420 pour l'estimation obtenue avec $S(1;0;1)k=1.4$; G010 pour l'estimation avec $A(0.5;0;1)k=1.5$. Dans les autres cas, elle est assez supérieure à sa valeur asymptotique. Dans la situation G210 avec l'estimateur $S(0.75;0.25;1)k=1.25$, elle est même 2.71 fois plus élevée que la valeur asymptotique. Par contre, la variance de $\hat{\mu}_n$ semble être stabilisée pour $n=50$ car les différences avec les valeurs trouvées pour $n=100$ sont

en général faibles.

Un autre phénomène est remarquable : la variance de $\tilde{\mu}_n$ reste nettement plus élevée que la valeur asymptotique pour $n=100$ dans les situations G410 et G420 avec les estimateurs $A(0.5;0;1)k=1.5$, $S(1;0;1)k=1.0$ et $S(0.75;0.25;1)k=1.25$. On s'aperçoit en fait que ces situations font partie des cas où μ_c , valeur vers laquelle converge $\tilde{\mu}_n$, est sensiblement différente de μ , moyenne du processus contaminé. Ce fait nous amène à douter de la généralité de la formule de la variance asymptotique de $\tilde{\mu}_n$ donnée par Martin (1981) et le problème de ses conditions de validité reste ouvert.

Remarquons également, fait curieux, que la variance de $\tilde{\mu}_n$ pour l'estimateur $S(1;0;1)k=1.4$ est supérieure à celle obtenue par les moindres carrés, dans la situation G420 et pour les 3 longueurs d'échantillons considérées.

En ce qui concerne le paramètre θ , on remarque que quel que soit l'estimateur considéré, il est biaisé vers le bas mais convergent, ce qui est un résultat attendu.

Le biais trouvé avec l'estimateur des moindres carrés correspond à l'approximation au premier ordre donné en I.3. Toutefois, avec les autres estimateurs le biais est quelque peu réduit pour les échantillons de taille 20, particulièrement avec l'estimateur

$A(0.5;0;1)k=1.5$ pour les contaminations G410 et G420.

Les variances des estimateurs de θ sont proches des valeurs asymptotiques déjà pour les échantillons de taille 50 sauf pour l'estimateur $S(0.75;0.25;1)k=1.25$ dans la situation G010.

Donc d'une manière générale une taille d'échantillon de 50 est suffisante pour pouvoir appliquer les formules asymptotiques avec la restriction déjà mentionnée concernant la variance des estimateurs de μ_v lorsque $\mu_c \neq \mu$.

Comparaisons des estimateurs

Quand la contamination est nulle, l'estimateur des moindres carrés est le meilleur. Il estime μ_v sans biais et la variance de $\hat{\mu}$ et $\hat{\theta}$ est minimale. Dans cette situation, l'estimateur asymétrique est le plus mauvais : $\tilde{\mu}_n$ a un biais important et l'EQM de $\tilde{\mu}_n$ est plus élevée que celui des autres estimateurs.

Lorsque la contamination est symétrique le comportement de l'estimateur asymétrique est encore plus mauvais. Le biais de $\tilde{\mu}_n$ augmente et sa variance est environ 2 fois plus élevée que celle qu'on obtient avec les estimateurs $S(1;0;1)k=1.4$ et $S(0.75;0.25;1)k=1.25$ et elle est également supérieure à celle que donne les moindres carrés. Les résultats des simulations pour $S(1;0;1)k=1.4$ et $S(0.75;0.25;1)k=1.25$ sont à peu près égaux dans cette

situation, mais les résultats asymptotiques avantagent quelque peu l'estimateur $S(0.75;0.25;1)k=1.25$.

Lorsque la contamination est faible et asymétrique, l'estimateur asymétrique reste peu approprié pour $\hat{\mu}_n$. Sa variance est alors plus élevée que celle obtenue avec les 3 autres estimateurs et son biais est plus grand que celui des estimateurs $S(1;0;1)k=1.4$ et $S(0.75;0.25;1)k=1.25$. Pour cette situation, les résultats des simulations feront préférer l'estimateur $S(1;0;1)k=1.4$.

Dans la situation G410 représentant une contamination asymétrique plus accentuée, l'estimateur asymétrique est alors intéressant. En valeur absolue, le biais de $\hat{\mu}_n$ est plus petit que celui qui résulte des 3 autres estimateurs. De plus pour $n=100$ la variance de $\hat{\mu}_n$ est inférieure aux variances trouvées avec les estimateurs $S(1;0;1)k=1.4$ et les moindres carrés. Seul l'estimateur $S(0.75;0.25;1)k=1.25$ a une variance inférieure mais un EQM supérieur.

Un autre avantage de l'estimateur asymétrique est que la variance de θ_n est inférieure à celles que donnent les autres estimateurs. De plus, toutes les valeurs asymptotiques favorisent cet estimateur.

Enfin, lorsque la contamination est forte et asymétrique (G420), l'estimateur asymétrique accentue ses qualités : le biais et la variance de $\hat{\mu}_n$ sont inférieurs à ceux que l'on trouve avec les 3 autres es-

estimateurs. De plus, la variance de $\hat{\theta}_n$ est également inférieure aux 3 autres variances.

En résumé, on retrouve les conclusions de la première partie concernant le choix des estimateurs selon les situations de contamination, à savoir :

- l'estimateur des moindres carrés est le meilleur possible lorsque les erreurs ne sont pas contaminées.
- Lorsque les erreurs sont contaminées de manière symétrique, les estimateurs $S(1;0;1)k=1.4$ et $S(0.75;0.25;1)k=1.25$ donnent de bons résultats. Les variances asymptotiques du second estimateur sont toutefois plus faibles. Il est donc à recommander.
- En cas de contamination légèrement asymétrique des erreurs (souvent invisible en pratique) l'estimateur $S(0.75;0.25;1)k=1.25$ est celui pour lequel le biais de $\hat{\mu}_n$ est le plus petit. Par contre, les variances sont légèrement plus élevées que celles obtenues avec $S(1;0;1)k=1.4$ et les moindres carrés, et donc, du point de vue de la minimisation de l'EQM, on préférera l'estimateur $S(1;0;1)k=1.4$.
- Lorsque la contamination asymétrique est accentuée, l'estimateur $A(0.5;0;1)k=1.5$ est à recommander car alors le biais de $\hat{\mu}_n$ est inférieur à celui que restituent les autres estimateurs et les variances de

$\hat{\mu}_n$ et $\hat{\theta}_n$ sont également inférieures aux variances que donnent les autres estimateurs.

Chapitre III. TRAITEMENT D'UN EXEMPLE

A titre d'exemple, nous avons repris une série se trouvant dans Kendall (1976, p. 17). Il s'agit d'une série trimestrielle allant du premier trimestre 1960 au quatrième trimestre 1971, formée d'un indice des actions cotées à la bourse de Londres et calculée par le Financial Times.

Afin de rendre la contamination plus apparente 4 valeurs sur les 48 ont été modifiées. Il s'agit des observations suivantes :

<u>Date de l'observation</u>	<u>No d'ordre de l'observation</u>	<u>valeur observée</u>	<u>valeur modifiée</u>
3ème trim.1961	7	310.4	280.0
3ème trim.1966	27	320.0	280.0
2ème trim.1968	34	461.1	435.0
3ème trim.1968	35	491.4	465.0

Cette série ainsi modifiée est dessinée sur la figure III.1.1.

Afin de savoir si cette série peut être modélisée par un processus autorégressif, on étudie les fonctions d'autocorrélation et d'autocorrélation partielle. Ces fonctions sont reproduites sur les figures III.1.3 et III.1.4, respectivement. La fonction d'autocorrélation descend linéairement jusqu'à un écart de 9 termes pour remonter légèrement aux écarts

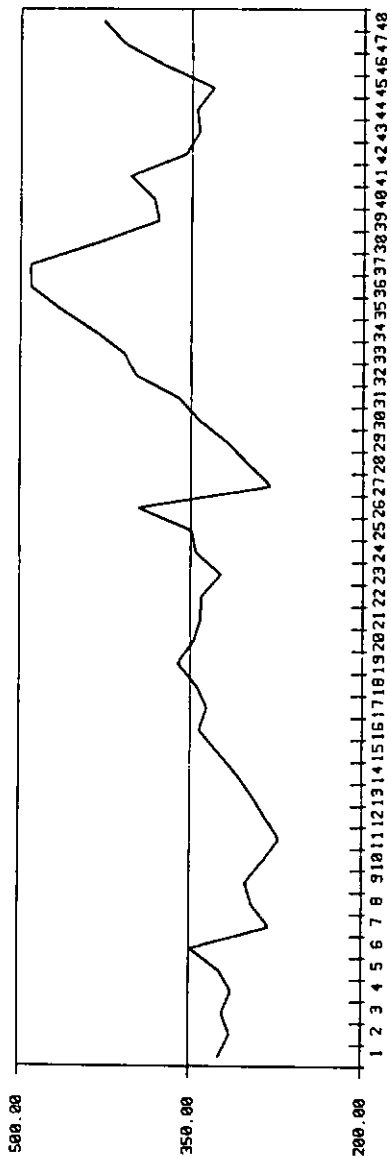


Figure III.1.1

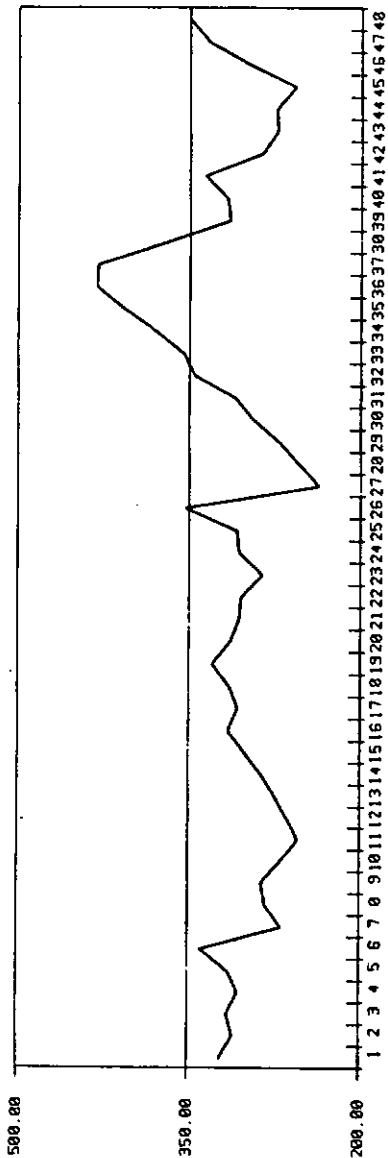


Figure III.1.2

10 et 11 et se stabiliser par la suite. Cette diminution linéaire est typique des processus autorégressifs admettant une tendance systématique. La fonction d'autocorrélation partielle descend brusquement à l'écart 2, ce qui est également une caractéristique des processus autorégressifs d'ordre 1.

En observant la figure III.1.1, on aperçoit une faible tendance systématique qu'il faut estimer et éliminer de la série afin d'obtenir une série stationnaire. Plusieurs méthodes ont été utilisées : moyennes mobiles, lissages robustes, trends linéaires.

Les moyennes mobiles qui exprimaient la tendance systématique sans trap de variations comprenaient entre 15 et 20 termes. Comme la série observée est de 48 termes, la série obtenue sans trend aurait été trop courte pour être étudiée. Les lissages robustes utilisés sont ceux qui sont proposés dans le package MINITAB. Les résultats n'étaient pas satisfaisants car le trend estimé était trop fluctuant. Nous avons donc finalement estimé la tendance par un ajustement linéaire. L'accroissement estimé de la valeur de l'indice est de 1.62 par trimestre et son écart type vaut 0.48. La série des indices sans trend a été dessinée sur la figure III.1.2. C'est cette série que nous allons étudier.

Les figures III.1.5 et III.1.6 représentent les fonctions d'autocorrélation et d'autocorrélation

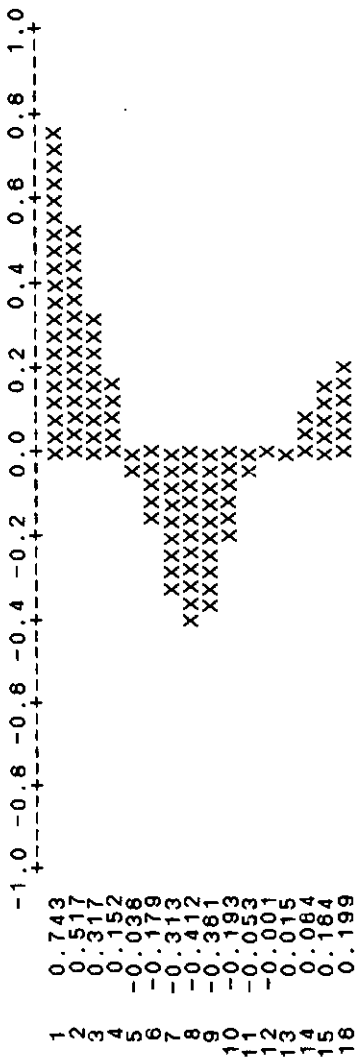


Figure III.1.5

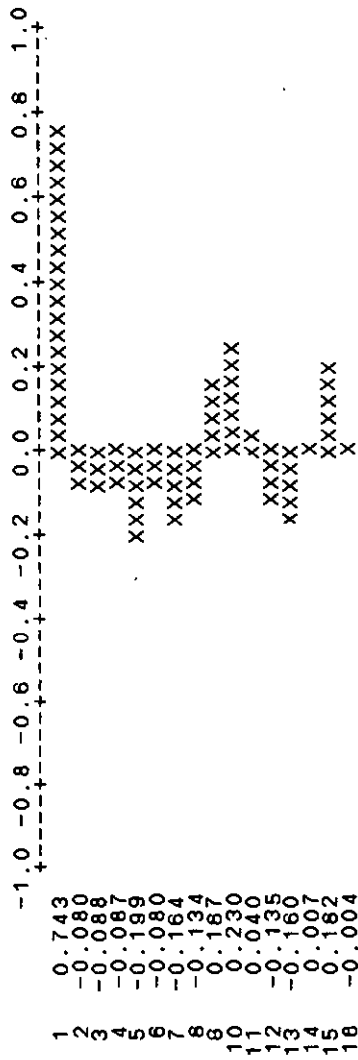


Figure III.1.6

partielle de la série démunie du trend. Cette fois la fonction d'autocorrélation a la forme d'une sinusoïde amortie, forme que l'on retrouve dans les processus autorégressifs. Comme aucun des coefficients d'autocorrélation partielle d'écart supérieur à 1 est significativement différent de 0, on peut en conclure que le modèle représentant correctement cette série est le processus AR(1). Les confirmations ont été apportées en ajustant des modèles AR(p) $p \leq 4$ avec MINITAB. Mis à part pour θ_1 , aucune autre estimation des θ_i étoit significativement différente de 0 pour les 4 modèles. L'estimation par un modèle AR(1) a donné les résultats suivants (estimation par la méthode des moindres carrés) :

	vol. estimée	Ecart-type	rapport-T
θ_1	0.757	0.098	7.71
δ	76.505	4.143	18.47

La valeur estimée de μ est donc 314.835.

Si on dessine le graphe des couples (x_{i-1}, x_i) $i=2, \dots, n$, on obtient la figure III.1.7. En observant cette figure, on remarque que 4 couples se détachent nettement des autres. Il s'agit des couples (x_6, x_7) , (x_{26}, x_{27}) , (x_{37}, x_{38}) , (x_{38}, x_{39}) . Un cinquième couple (x_{41}, x_{42}) se trouve également assez éloigné des autres observations. Tous ces couples correspondent à des chutes brusques de la valeur de

l'indice observé, chutes provoquées par les modifications des voleurs x_6 et x_{27} dans les 2 premiers cas. Comme on représente la série par un modèle AR(1), on peut considérer que ces brusques chutes sont dues à une grande erreur négative aux points x_6 , x_{26} , x_{37} , x_{38} et x_{41} . On en déduit donc que la loi des erreurs U_t n'est pas symétrique autour de 0 mais asymétrique avec une queue allongée à gauche. Afin de réduire l'influence de ces observations, on va estimer les paramètres du modèle AR(1) en utilisant des estimateurs robustes. Pour différents estimateurs les résultats obtenus sont les suivants :

Estimateur	val. estimée de θ_1	val. estimée de μ
$S(1;0;1)k=1.4$	0.810	327.363
$S(0.75;0.25;1)k=1.25$	0.831	337.691
$A(0.5;0;1)k=1.5$	0.883	370.367

La figure III.1.8 représente les droites de régression correspondant aux paramètres estimés respectivement par les estimateurs des moindres carrés (1), $S(1;0;1)k=1.4$ (2) $S(0.75;0.25;1)k=1.25$ (3) et $A(0.5;0;1)k=1.5$ (4). Les figures III.1.9, 10, 11 et 12 représentent l'indice observé en trait gros et l'indice estimé en trait fin par chacun de ces 4 estimateurs.

Il est évidemment impossible de dire quelle

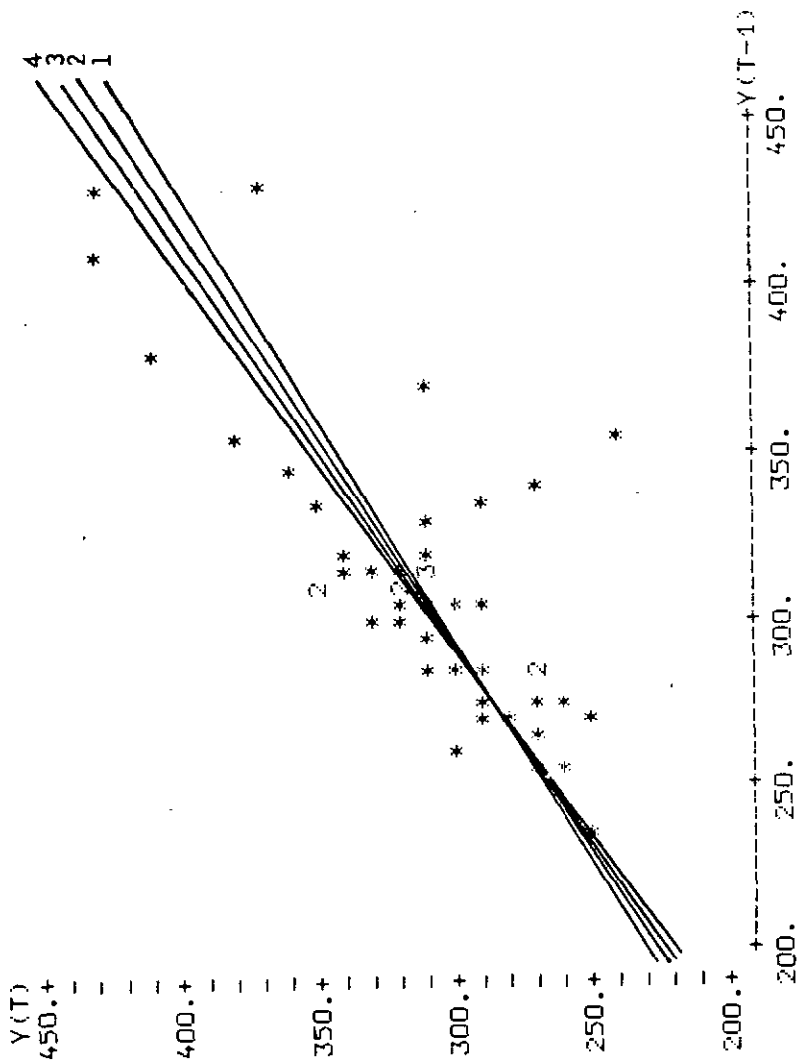


Figure III.1.1.8

500.00

350.00

200.00

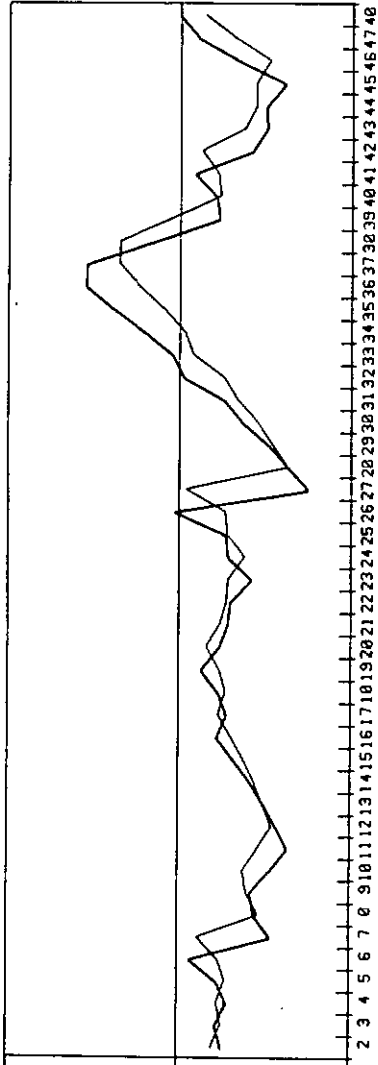


Figure III.1.9

500.00

350.00

200.00

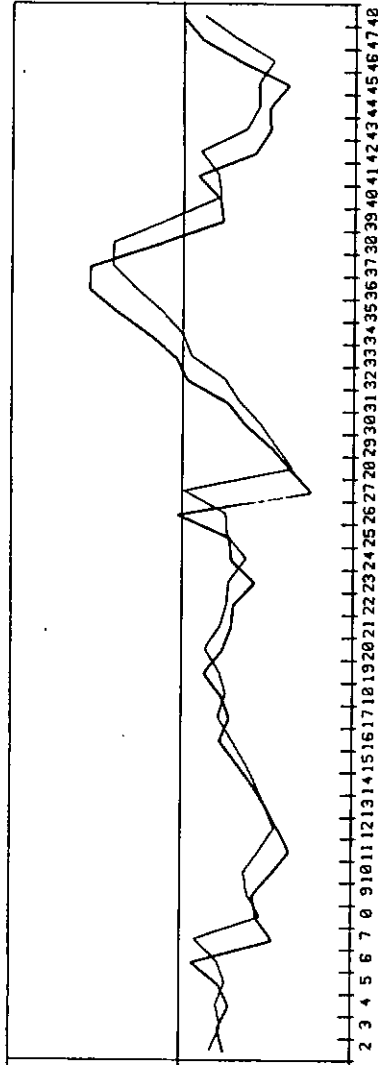


Figure III.1.10

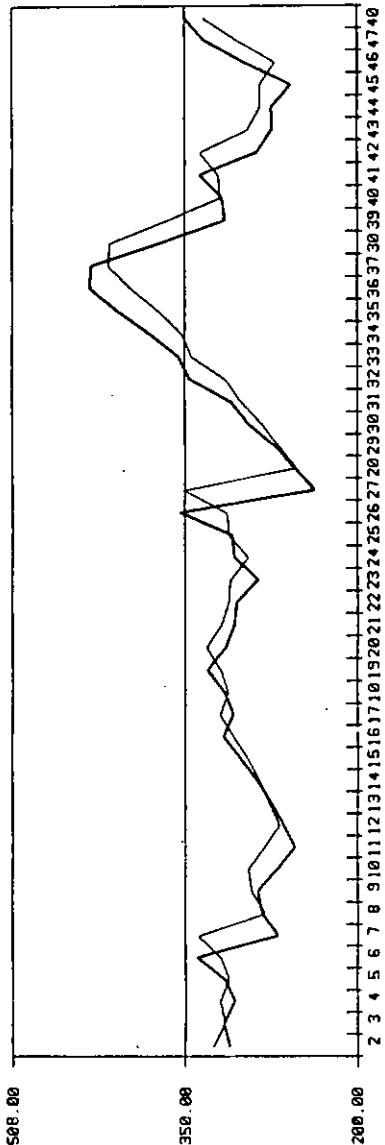


Figure III.1.1.11

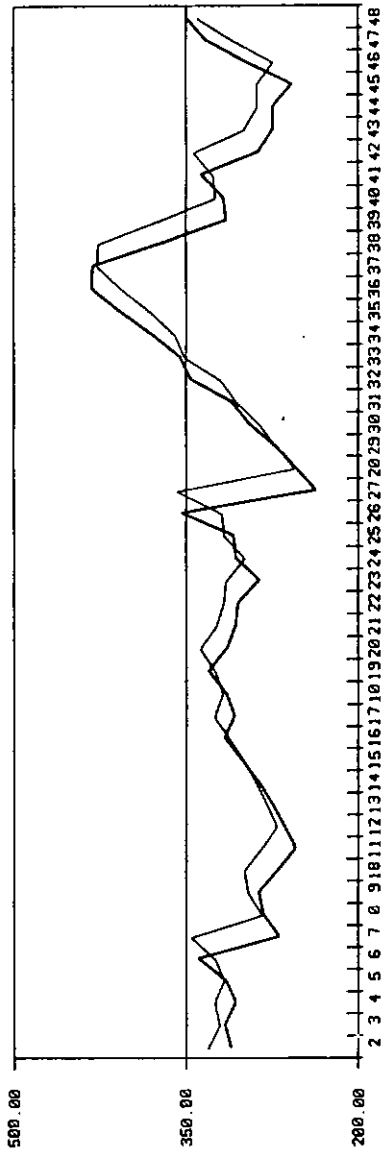


Figure III.1.1.12

est la meilleure estimation puisque le modèle idéal est inconnu. Si, par exemple, nous décidons que la meilleure estimation est celle qui minimise la somme des écarts entre les valeurs observées et les valeurs calculées élevées au carré, nous privilégions l'estimation des moindres carrés qui justement est définie par ce critère.

Toutefois, en examinant les résidus, il est possible de proposer une estimation convenable. Pour cela, nous partons de la méthode des moindres carrés. Nous savons que cette méthode tend à symétriser les résidus. Si, malgré cet effet, les résidus sont répartis de manière asymétrique, on en déduit, du point de vue de l'estimation robuste, que le modèle idéal a été contaminé de façon asymétrique.

Dans notre cas, nous avons estimé la densité des résidus provenant de la régression par les moindres carrés avec une méthode d'estimation de densité à noyau variable implantée dans le programme RESISTA, écrit par M. Lejeune et G. Maitre, GRMQ, Université de Neuchâtel. On a reproduit cette densité estimée des résidus sur la figure III.1.13. On constate que la densité présente une asymétrie sous forme de queue allongée à gauche. Cela signifie que certains résidus sont très négatifs, et donc que le modèle est contaminé de manière asymétrique.

Quelle est l'importance de la contamination ?

Nous ne le savons évidemment pas avec exactitude. Nous pouvons toutefois essayer de l'estimer en se ramenant à des situations théoriques étudiées plus haut ou, du moins, à des situations proches.

Etant donné la présence d'une contamination asymétrique, nous obtiendrons une meilleure estimation des paramètres μ_v et θ et donc des résidus en utilisant l'estimateur $S(1;0;1)k=1.4$ plutôt que celui des moindres carrés. La courbe en trait épais sur la figure III.1.14 est une estimation de la densité de ces résidus. Nous avons essayé d'ajuster visuellement cette densité estimée à une densité connue de la forme $(1-\epsilon) (\mu_1, \sigma_1) + \epsilon (\mu_2, \sigma_2)$, qui est la densité dessinée sur la même figure, en trait fin, avec les paramètres $\mu_1=0$, $\mu_2=-66$, $\sigma_1=\sigma_2=22$ et $\epsilon=0.05$. Donc, au signe près de la contamination et à un facteur d'échelle près cette densité représente une situation de contamination G305.

On peut faire le même raisonnement avec les résultats obtenus par régression avec l'estimateur $A(0.5;0;1)k=1.5$ (dont on connaît les qualités en situation de contamination asymétrique) qui sont assez différents des résultats obtenus avec l'estimateur $S(1;0;1)k=1.4$. A nouveau la densité estimée des résidus ainsi que son ajustement ont été dessinés sur la figure III.1.15. Les paramètres utilisés pour ajuster la densité des résidus sont $\mu_1=0$, $\mu_2=-65$, $\sigma_1=20$, $\sigma_2=23$ et $\epsilon=0.10$, c'est-à-dire au signe près de la contamination et à un facteur d'échelle près, à peu près

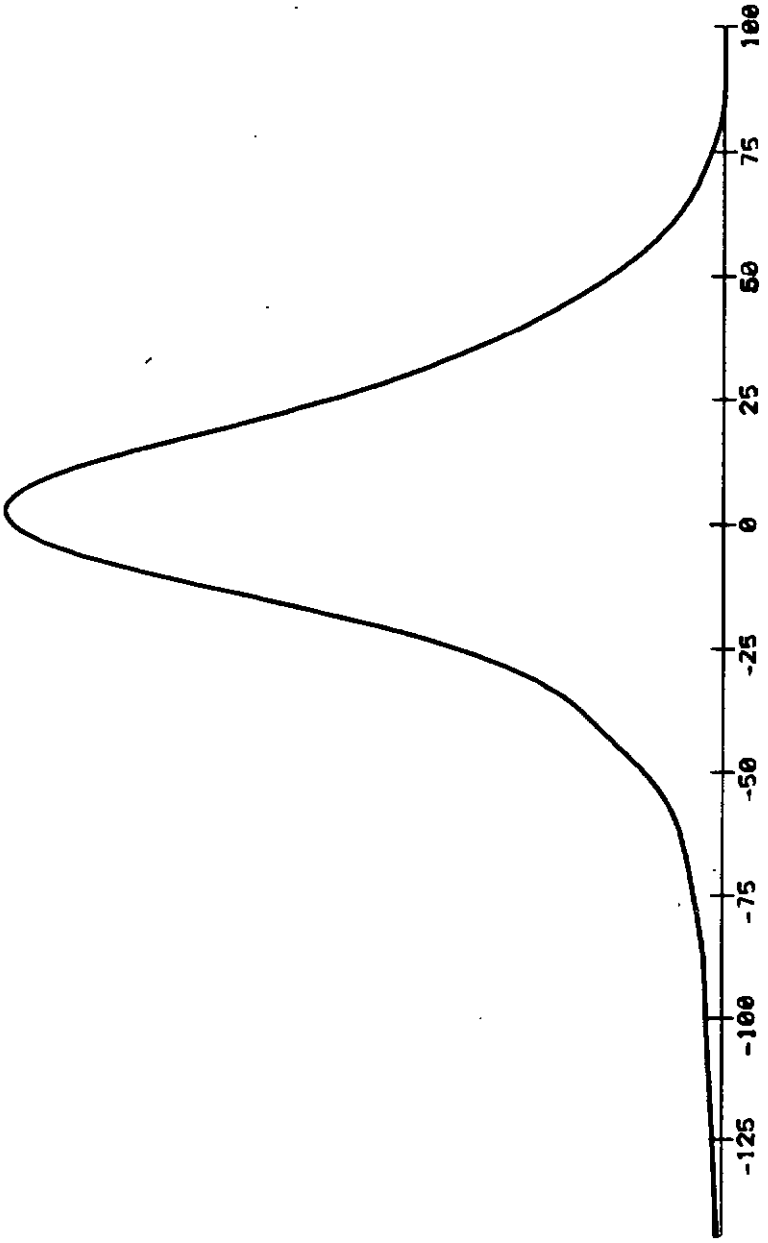


Figure III.1.13

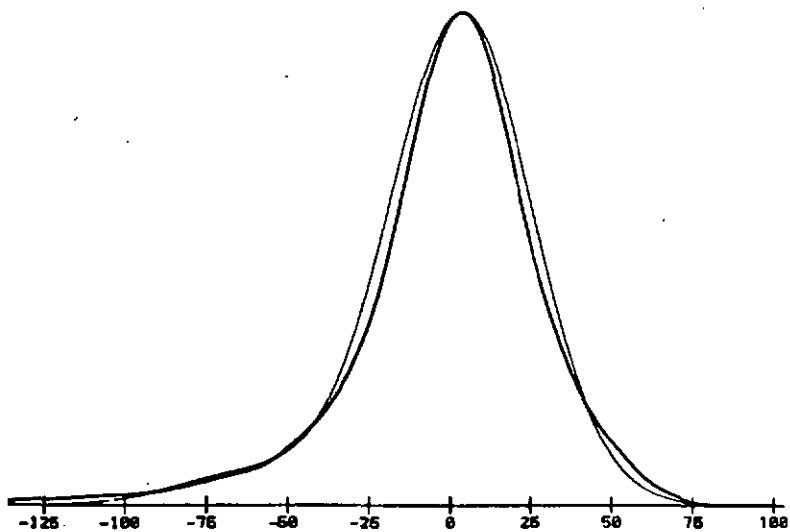


Figure III.1.14

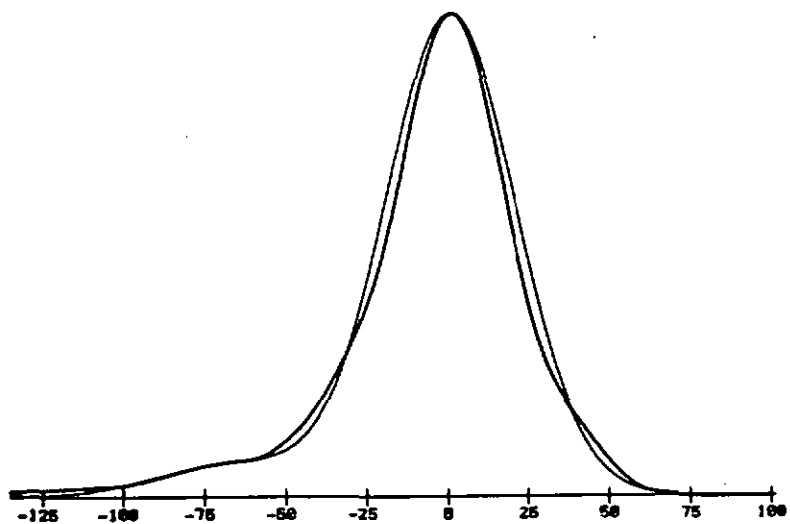


Figure III.1.15

une situation G310.

Il semble donc raisonnable de dire que la série étudiée est contaminée de manière asymétrique, avec une loi des erreurs se situant entre G305 et G310.

Nous nous trouvons ainsi dans un cas de contamination asymétrique moyenne, une sorte de "zone grise" pour laquelle il est difficile de recommander à priori un estimateur symétrique ou asymétrique. On s'en rend compte en examinant les résultats de simulation pour les contaminations G210 et G410, qui sont les cas entre lesquels nous nous trouvons. En se fiant aux résultats asymptotiques nous obtenons, pour les valeurs des paramètres $\mu_V=0.0$; $\theta=0.8$ et pour les contaminations G305 et G310, les résultats suivants, pour $\hat{\mu}$:

		G305	G310
S(1;0;1)k=∞	MOY $\hat{\mu}$	0.749	1.499
	VAR $\hat{\mu}$	35.687	45.250
S(1;0;1)k=1.4	MOY $\hat{\mu}$	0.471	1.056
	VAR $\hat{\mu}$	31.322	39.298
S(0.75;0.25;1)k=1.25	MOY $\hat{\mu}$	0.407	0.895
	VAR $\hat{\mu}$	32.413	38.770
A(0.5;0;1)k=1.5	MOY $\hat{\mu}$	-1.140	-0.656
	VAR $\hat{\mu}$	30.584	35.207

Si la contamination est de la forme G305, les 3 derniers estimateurs sont à peu près équivalents du point de vue de l'EQM. Par contre, si la contamination peut être modélisée par G310, il convient d'utiliser l'estimateur $A(0.5;0;1)_{k=1.5}$ car son EQM est nettement inférieur aux 3 autres.

Comme nous désirons minimiser l'EQM, nous retiendrons pour cet exemple les estimations des paramètres θ et μ_v obtenues en utilisant l'estimateur asymétrique $A(0.5;0;1)_{k=1.5}$.

ANNEXE I

Calcul de la variance asymptotique des estimateurs de la moyenne d'une loi normale contaminée

La variance asymptotique vaut :

$$\frac{\int \psi^2 \left(\frac{x-c}{\sigma_x} \right) dF(x)}{(\lambda'(c))^2}$$

$$\text{où } \lambda(u) = \int \psi \left(\frac{x-u}{\sigma_x} \right) dF(x)$$

$$\text{et } \lambda'(c) = \left[\frac{d}{du} \lambda(u) \right]_{u=c}$$

où c est défini implicitement par $\lambda(c)=0$

Si $f(x) = \frac{d}{dx} F(x)$ est dérivable, alors

$$\lambda'(c) = \int \psi \left(\frac{x-c}{\sigma_x} \right) f'(x) dx$$

La suite des opérations à effectuer est la suivante :

- i) calculer σ_x
- ii) calculer la valeur $c = \lim_{n \rightarrow \infty} E(T_n)$, qui représente également le biais de l'estimateur
- iii) calculer la variance asymptotique de l'estimateur.

Nous ne donnons ici le développement des calculs que pour les fonctions ψ asymétriques, les calculs concernant les fonctions ψ symétriques se font d'une manière analogue.

La fonction ψ est définie comme :

$$\psi(u) = \begin{cases} u & \text{si } u \leq 0 \\ \alpha u + \beta k & \text{si } 0 < u \leq k \\ \gamma k(\alpha + \beta) & \text{si } u > k \end{cases}$$

où $u = \frac{x-c}{\sigma_x}$ et $c = \lim_{n \rightarrow \infty} E(T_n)$

o) Pour une combinaison convexe de deux lois normales

La densité de probabilité est définie comme :

$$f(x) = (1-\epsilon)\varphi(x) + \epsilon\varphi_1(x)$$

où $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} x^2)$

$$\varphi_1(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2} (\frac{x-m}{\sigma})^2)$$

i) $\sigma_x^2 = (1-\epsilon)(1+m^2\epsilon) + \epsilon\sigma^2$

ii) $\lim_{n \rightarrow \infty} E(T_n) = c \Leftrightarrow \int \psi(\frac{x-c}{\sigma_x}) f(x) dx = 0$

On obtient

$$\int_{-\infty}^c (\frac{x-c}{\sigma_x}) f(x) dx + \frac{\alpha}{\sigma_x} \int_c^{c+k\sigma_x} (x-c) f(x) dx +$$

$$+ \beta k \int_c^{c+k\sigma} x f(x) dx + \gamma k(\alpha+\beta) \int_{c+k\sigma}^{\infty} f(x) dx = 0$$

$$\text{soit, en posant } B = c+k\sigma_x ; B' = \frac{c+k\sigma_x^{-m}}{\sigma} ;$$

$$c' = \frac{c-m}{\sigma} \text{ et } \phi(x) = \int_{-\infty}^x \varphi(u) du$$

et après développement,

$$- \left(\frac{1-\varepsilon}{\sigma_x} \right) (\varphi(c) + c\phi(c)) + \frac{\varepsilon}{\sigma_x} (m\phi(c') - \sigma\varphi(c')) - c\phi(c')$$

$$+ \frac{\alpha}{\sigma_x} \{ (1-\varepsilon)[- \varphi(B) + \varphi(c) - c(\phi(B) - \phi(c))] +$$

$$+ \varepsilon[m(\phi(B') - \phi(c')) + \sigma(\varphi(c') - \varphi(B')) - c(\phi(B') - \phi(c'))] \}$$

$$+ \beta k[(1-\varepsilon)(\phi(B) - \phi(c)) + \varepsilon(\phi(B') - \phi(c'))]$$

$$+ \gamma k(\alpha+\beta)[(1-\varepsilon)(1-\phi(B)) + \varepsilon(1-\phi(B'))] = 0$$

La solution c de cette équation a été obtenue par un algorithme itératif.

iii) Calcul de la variance asymptotique

Nous procédons d'abord au calcul du numérateur

$$\int_{-\infty}^{+\infty} \psi^2(u) f(x) dx = \frac{1}{\sigma_x^2} \int_{-\infty}^c (x-c)^2 f(x) dx +$$

$$+ \frac{\alpha}{\sigma_x^2} \int_c^{c+k\sigma} x (x-c)^2 f(x) dx + \frac{2\beta k\alpha}{\sigma_x} \int_c^{c+k\sigma} x (x-c) f(x) dx +$$

$$+ k^2 \beta^2 \int_c^{c+k\sigma_x} f(x) dx + (\gamma k(\alpha+\beta))^2 \int_{c+k\sigma_x}^{+\infty} f(x) dx$$

Nous obtenons, oprès développement :

$$\begin{aligned} & \frac{1-\varepsilon}{\sigma_x^2} (-c\varphi(c) + \phi(c)) + \frac{\varepsilon}{\sigma_x^2} [\sigma^2 (-c'\varphi(c') + \phi(c')) + \\ & + 2\mu\sigma(-\varphi(c')) + \mu^2\phi(c')] + \frac{1}{\sigma_x^2} [2c\varphi(c)(1-\varepsilon) - \\ & - 2c\varepsilon(\mu\phi(c') - \sigma\varphi(c')) + (1-\varepsilon)c^2\phi(c) + c^2\varepsilon\phi(c')] + \\ & + \frac{\alpha}{\sigma_x^2} \{ (1-\varepsilon)(c\varphi(c) - B\varphi(B) + \phi(B) - \phi(c)) + \\ & + \varepsilon(\sigma^2(-B'\varphi(B') + c'\varphi(c') + \phi(B') - \phi(c')) + \\ & + 2m\sigma(\varphi(c') - \varphi(B')) + m^2(\phi(B') - \phi(c'))) - \\ & - 2c(1-\varepsilon)(\varphi(c) - \varphi(B)) - 2c\varepsilon(m(\phi(B') - \phi(c')) - \\ & - \sigma(\varphi(B') - \varphi(c'))) + c^2(1-\varepsilon)(\phi(B) - \phi(c)) + \\ & + c^2\varepsilon(\phi(B') - \phi(c')) \} + \\ & + \frac{2\beta k\alpha}{\sigma_x} \{ (1-\varepsilon)(\varphi(c) - \varphi(B)) + \varepsilon[\mu(\phi(B') - \phi(c')) - \\ & - \sigma(\varphi(B') - \varphi(c'))] - c(1-\varepsilon)(\phi(B) - \phi(c)) + \end{aligned}$$

$$\begin{aligned}
& + \epsilon c (\phi(B') - \phi(c')) + \\
& + \beta^2 k^2 (1-\epsilon) (\phi(B) - \phi(c)) + \beta^2 k^2 \epsilon (\phi(B') - \phi(c')) + \\
& + (\gamma k (\alpha + \beta))^2 [(1-\epsilon)(1-\phi(B) + \epsilon(1-\phi(B')))]
\end{aligned}$$

Le dénominateur vaut $(\lambda'(u))_{u=c}^2$.

$$\begin{aligned}
\text{Or } \lambda(u) &= \int \psi\left(\frac{x-u}{\sigma}\right) f(x) dx \\
&= \frac{1}{\sigma_x} \int_{-\infty}^u (x-u) f(x) dx + \frac{\alpha}{\sigma_x} \int_u^{u+k\sigma_x} (x-u) f(x) dx + \\
&+ \beta k \int_u^{u+k\sigma_x} f(x) dx + \gamma k (\alpha + \beta) \int_{u+k\sigma_x}^{+\infty} f(x) dx
\end{aligned}$$

En dérivant sous le signe somme on obtient pour $u = c$:

$$\begin{aligned}
\lambda'(u)_{u=c} &= -\frac{1}{\sigma_x} F(c) + \frac{\alpha}{\sigma_x} [k\sigma_x f(B) - F(B) + F(c)] + \\
&+ \beta k (f(B) - f(c)) - \gamma k (\alpha + \beta) f(B)
\end{aligned}$$

soit, en substituant f par sa valeur

$$\begin{aligned}
\lambda'(u)_{u=c} &= -\frac{1}{\sigma_x} [(1-\epsilon)\phi(c) + \epsilon\phi(c')] + \\
&+ \frac{\alpha}{\sigma_x} \{k\sigma_x [(1-\epsilon)\phi(B) + \epsilon\phi(B')] -
\end{aligned}$$

$$\begin{aligned}
& - [(1-\varepsilon)\phi(B) + \varepsilon\phi(B')] + \\
& + (1-\varepsilon)\phi(c) + \varepsilon\phi(c') - \\
& - c[(1-\varepsilon)\phi(B) + \varepsilon\phi(B')] + \\
& + \beta k[(1-\varepsilon)\phi(B) + \varepsilon\phi(B')] - \\
& - (1-\varepsilon)\phi(c) - \varepsilon\phi(c') + \\
& + \gamma k(\alpha + \beta)[(1-\varepsilon)\phi(B) + \varepsilon\phi(B')]
\end{aligned}$$

b) Pour une loi normale contaminée par une loi exponentielle

La densité de probabilité est définie comme :

$$\begin{aligned}
f(x) &= (1-\varepsilon)\varphi(x) + \mu e^{-\lambda(x-a)} I_a(x) - \\
& - (1-\varepsilon)\varphi(x) I_0(x)
\end{aligned}$$

où, comme auparavant,

$\varphi(x)$ est la fonction de densité d'une loi de

Gauss centrée et réduite

$I_0(x)$ est la fonction indicatrice qui vaut

1 si $x \in]a, +\infty[$ et 0 sinon,

$$\mu = (1-\varepsilon)\varphi(a) ,$$

$$\lambda = \frac{\mu}{1-(1-\varepsilon)\varphi(a)} ,$$

$$\phi(x) = \int_{-\infty}^x \varphi(u) du .$$

$$i) \quad E(x) = \frac{\mu}{\lambda^2} (a\lambda + 1 - \lambda^2)$$

$$V(x) = \sigma_x^2 = 1 - \frac{\mu}{\lambda} - a\mu + \frac{\mu}{\lambda^3} (a^2\lambda^2 + 2a\lambda + 2) - (E(x))^2$$

$$ii) \quad \lim_{n \rightarrow \infty} E(Tn) = c \Leftrightarrow \int \psi\left(\frac{x-c}{\sigma_x}\right) f(x) dx = 0$$

$$\Leftrightarrow \lambda(c) = 0$$

Soit, en posant $cp = c + k\sigma_x$, $mc = \max(a, c)$,

$mck = \max(a, c + k\sigma_x)$ et $e(x) = e^{-\lambda(x-a)}$

$$\frac{(1-\varepsilon)}{\sigma_x} \int_{-\infty}^c x\varphi(x) dx + \frac{\mu}{\sigma_x} \int_a^{mc} xe(x) dx -$$

$$- \frac{(1-\varepsilon)}{\sigma_x} \int_a^{mc} x\varphi(x) dx - \frac{c}{\sigma_x} \left[(1-\varepsilon) \int_{-\infty}^c \varphi(x) dx + \right.$$

$$\left. + \mu \int_a^{mc} e(x) dx - (1-\varepsilon) \int_a^{mc} \varphi(x) dx \right] +$$

$$+ \frac{\alpha}{\sigma_x} \left\{ (1-\varepsilon) \int_c^{cp} x\varphi(x) dx + \mu \int_{mc}^{mck} xe(x) dx - \right.$$

$$\left. - (1-\varepsilon) \int_{mc}^{mck} x\varphi(x) dx - c \left[(1-\varepsilon) \int_c^{cp} \varphi(x) dx + \right. \right.$$

$$\begin{aligned}
& + \mu \int_{mc}^{mck} e(x) dx - (1-\varepsilon) \int_{mc}^{mck} \varphi(x) dx \} + \\
& + \beta k [(1-\varepsilon) \int_c^{cp} \varphi(x) dx + \mu \int_{mc}^{mck} e(x) dx - (1-\varepsilon) \int_{mc}^{mck} \varphi(x) dx] + \\
& + \gamma k (\alpha + \beta) [(1-\varepsilon) \int_{cp}^{+\infty} \varphi(x) dx + \mu \int_{mck}^{+\infty} e(x) dx - \\
& - (1-\varepsilon) \int_{mck}^{+\infty} \varphi(x) dx] = 0
\end{aligned}$$

En utilisant la fonction $I_a(x)$, on obtient après développement et simplifications :

$$\begin{aligned}
& - \frac{(1-\varepsilon)}{\sigma_x} \varphi(c) + I_a(c) \frac{\mu}{\sigma_x \lambda} (-ce(c) + \alpha - \frac{e(c)}{\lambda} + \frac{1}{\lambda}) - \\
& - \frac{(1-\varepsilon)}{\sigma_x} I_a(c) (-\varphi(c) + \varphi(\alpha)) - \frac{c}{\sigma_x} \{ (1-\varepsilon) \phi(c) + \\
& + I_0(c) [\frac{\mu}{\lambda} (1-e(c)) - (1-\varepsilon) (\phi(c) - \phi(\alpha))] \} + \\
& + \frac{\alpha}{\sigma_x} \{ (1-\varepsilon) (\varphi(c) - \varphi(cp)) + I_a(c) I_0(cp) \frac{\mu}{\lambda^2} (-e(cp) + \\
& + e(c)) + I_a(cp) (1 - I_a(c)) \frac{\mu}{\lambda} (-cpe(cp) + \alpha - \\
& - \frac{e(cp)}{\lambda} + \frac{1}{\lambda}) - I_a(c) I_a(cp) (1-\varepsilon) (\varphi(\alpha) - \varphi(cp)) - \\
& - I_a(cp) (1 - I_a(c)) (1-\varepsilon) (\varphi(\alpha) - \varphi(cp)) \} - \\
& - \frac{\alpha c}{\sigma_x} \{ (1-\varepsilon) (\phi(cp) - \phi(c)) + I_a(cp) (1 - I_a(c)) [\frac{\mu}{\lambda} (1-e(cp)) -
\end{aligned}$$

$$\begin{aligned}
& - (1-\epsilon)(\phi(cp)-\phi(a))] - \\
& - I_a(c)I_a(cp)(1-\epsilon)(\phi(cp)-\phi(c)) \} + \\
& + \beta k\{(1-\epsilon)(\phi(cp)-\phi(c))+I_a(c)\frac{\mu}{\lambda}(e(c)-e(cp)) + \\
& + I_a(cp)(1-I_a(c))[1-e(cp)-(1-\epsilon)(\phi(cp)-\phi(a))]\} - \\
& - I_a(c)(1-\epsilon)(\phi(cp)-\phi(c))\} \\
& + \gamma k(\alpha+\beta)\{(1-\epsilon)(1-\phi(cp)) + I_a(cp)[\frac{\mu}{\lambda}e(cp) - \\
& - (1-\epsilon)(1-\phi(cp))]\} + (1-I_a(cp))[\frac{\mu}{\lambda} - \\
& - (1-\epsilon)(1-\phi(a))]\} = 0
\end{aligned}$$

iii) Calcul de la variance asymptotique

Nous effectuons d'abord le calcul du numérateur qui vaut :

$$\int \psi^2 \left(\frac{x-c}{\sigma_x} \right) f(x) dx$$

Après développement de l'expression, on obtient :

$$\begin{aligned}
& \frac{(1-\varepsilon)}{\sigma_x^2} (-c\varphi(c) + \phi(c)) + \frac{\mu}{\sigma_x^2} \left[\frac{e(a)}{\lambda} \left(a^2 + \frac{2a}{\lambda} + \frac{2}{\lambda^2} \right) - \right. \\
& \left. - \frac{e(mc)}{\lambda} \left((mc)^2 + \frac{2mc}{\lambda} + \frac{2}{\lambda^2} \right) \right] - \\
& - \frac{(1-\varepsilon)}{\sigma_x^2} (a\varphi(a) - \phi(a) - mc\varphi(mc) + \phi(mc)) - \\
& - \frac{2c}{\sigma_x^2} [(1-\varepsilon)(-\varphi(c)) + \mu \left[\frac{e(a)a}{\lambda} + \frac{e(a)}{\lambda^2} - \frac{mc}{\lambda} e(mc) - \frac{1}{\lambda^2} e(mc) \right] - \\
& - (1-\varepsilon)(\varphi(a) - \varphi(mc))] + \frac{c^2}{\sigma_x^2} [(1-\varepsilon)\phi(c) + \\
& + \frac{\mu}{\lambda} (e(a) - e(mc)) - (1-\varepsilon)(\phi(mc) - \phi(a))] + \\
& + \frac{\alpha^2}{\sigma_x^2} [(1-\varepsilon)(c\varphi(c) - (cp)\varphi(cp) + \phi(cp) - \phi(c)) + \\
& + \mu \left[\frac{e(mc)}{\lambda} \left((mc)^2 + \frac{2mc}{\lambda} + \frac{2}{\lambda^2} \right) - \right. \\
& \left. - \frac{e(mck)}{\lambda} \left((mck)^2 + \frac{2mck}{\lambda} + \frac{2}{\lambda^2} \right) \right] - \\
& - (1-\varepsilon)(mc\varphi(mc) - mck\varphi(mck) + \phi(mck) - \phi(mc))] - \\
& - \frac{2c\alpha^2}{\sigma_x^2} [(1-\varepsilon)(\varphi(c) - \varphi(cp)) + \mu \left[\frac{mc}{\lambda} e(mc) + \frac{e(mc)}{\lambda^2} - \right. \\
& \left. - \frac{mck}{\lambda} e(mck) - \frac{e(mck)}{\lambda^2} \right] - (1-\varepsilon)(\varphi(mc) - \varphi(mck))] + \\
& + \frac{c^2\alpha^2}{\sigma_x^2} [(1-\varepsilon)(\phi(cp) - \phi(c)) + \frac{\mu}{\lambda} (e(mc) - e(mck))] -
\end{aligned}$$

$$\begin{aligned}
& - (1-\epsilon)(\phi(mck)-\phi(mc))] .+ \\
+ & \beta^2 k^2 [(1-\epsilon)(\phi(cp)-\phi(c)) + \frac{\mu}{\lambda}(e(mc)-e(mck)) - \\
& - (1-\epsilon)(\phi(mck)-\phi(mc))] \\
+ & (\gamma k(\alpha+\beta))^2 [(1-\epsilon)(1-\phi(cp)) + \frac{\mu}{\lambda} e(mck) - \\
& - (1-\epsilon)(1-\phi(mck))] + \frac{2\alpha\beta k}{\sigma_x} [(1-\epsilon)(\varphi(c) - \\
& - \varphi(cp)) + \frac{\mu}{\lambda}(e(mc)mc - e(mck)mck) + \frac{e(mc)}{\lambda} - \frac{e(mck)}{\lambda}) - \\
& - (1-\epsilon)(\varphi(mc)-\varphi(mck))] \\
- & \frac{2\alpha\beta kc}{\sigma_x} [(1-\epsilon)(\phi(cp)-\phi(c)) + \frac{\mu}{\lambda}(e(mc)-e(mck)) - \\
& - (1-\epsilon)(\phi(mck)-\phi(mc))]
\end{aligned}$$

Puisque $f(x)$ n'est pas dérivable en tout point, on ne peut dériver sous le signe somme et donc on calculera le dénominateur directement par

$$\left(\frac{d}{du} \lambda(u)\right)_{u=c}^2$$

Il suffit donc de dériver l'expression $\lambda(c)$, calculée en b.ii), par rapport à c et d'élever le résultat au carré. Après quelques simplifications, on

obtient le dénominateur suivant :

$$\begin{aligned}
 & - \frac{1}{\sigma_x} \{ (1-\varepsilon)\phi(c) + I_a(c) \left[\frac{\mu}{\lambda}(1-e(c)) - \right. \\
 & \quad \left. - (1-\varepsilon)(\phi(c) - \phi(a)) \right] \} + \\
 & + \frac{\alpha}{\sigma_x} \{ (1-\varepsilon)c_p\varphi(c_p) + I_a(c) [\mu c_p e(c_p) - (1-\varepsilon)c_p\varphi(c_p)] + \\
 & \quad + I_a(c_p)(1-I_a(c)) [\mu c_p e(c_p) - (1-\varepsilon)c_p\varphi(c_p)] \} - \\
 & - \frac{\alpha}{\sigma_x} \{ (1-\varepsilon)(\phi(c_p) - \phi(c)) + I_a(c) \left[\frac{\mu}{\lambda}(e(c) - e(c_p)) - \right. \\
 & \quad \left. - (1-\varepsilon)(\phi(c_p) - \phi(c)) \right] + I_a(c_p)(1-I_a(c)) \left[\frac{\mu}{\lambda}(1-e(c_p)) - \right. \\
 & \quad \left. - (1-\varepsilon)(\phi(c_p) - \phi(a)) \right] \} - \\
 & - \frac{\alpha c}{\sigma_x} \{ (1-\varepsilon)\varphi(c_p) + I_a(c) [\mu e(c_p) - (1-\varepsilon)\varphi(c_p)] + \\
 & \quad + I_a(c_p)(1-I_a(c)) [\mu e(c_p) - (1-\varepsilon)\varphi(c_p)] \} + \\
 & + \beta k \{ (1-\varepsilon)(\varphi(c_p) - \varphi(c)) + I_a(c) [\mu(e(c_p) - e(c)) - \\
 & \quad - (1-\varepsilon)(\varphi(c_p) - \varphi(c))] + I_a(c_p)(1-I_a(c)) [\mu e(c_p) - \\
 & \quad - (1-\varepsilon)\varphi(c_p)] \} + \\
 & + \gamma k (\alpha + \beta) \{ -(1-\varepsilon)\varphi(c_p) - I_a(c_p) [\mu e(c_p) - (1-\varepsilon)\varphi(c_p)] \}
 \end{aligned}$$

Calcul du EQM asymptotique des estimateurs de l'écart-type

1) Ecart semi-interquartile

a) Modèle de contamination gaussien

$$f(x) = (1-\varepsilon)\varphi_0(x; m_0, \sigma_0) + \varepsilon\varphi_1(x; m_1, \sigma_1)$$

$$\varphi_i(x; m_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x-m_i}{\sigma_i}\right)^2\right\}$$

$$\lim_{n \rightarrow \infty} E(T_n) = \frac{1}{2 \cdot 0.6745} \left(\xi_3 - \xi_1 \right)$$

$$\lim_{n \rightarrow \infty} \text{biais}(T_n) = \lim_{n \rightarrow \infty} E(T_n) - \sigma_0$$

$$\lim_{n \rightarrow \infty} V(\sqrt{n} T_n) = \frac{1}{64 \cdot (0.6745)^2} \cdot \left(\frac{3}{f^2(\xi_1)} + \frac{3}{f^2(\xi_3)} - \frac{2}{f(\xi_1)f(\xi_3)} \right)$$

$$\lim_{n \rightarrow \infty} \text{EQM}(T_n) = \lim_{n \rightarrow \infty} V(\sqrt{n} T_n) + (\lim_{n \rightarrow \infty} \text{biais}(T_n))^2$$

b) Modèle de contamination exponentiel

$$f(x) = (1-\varepsilon)\varphi_0(x; 0, 1) + \mu e^{-\lambda(x-a)} I_a(x) - (1-\varepsilon)\varphi_0(x; 0, 1) I_a(x)$$

L'expression est identique à 1.o) .

2) Moindres carrés

a) Modèle de contamination gaussien

$$\lim_{n \rightarrow \infty} E(T_n) = \sqrt{\mu_2}$$

$$\lim_{n \rightarrow \infty} \text{biais}(T_n) = \sqrt{\mu_2} - \sigma_0; \quad \lim_{n \rightarrow \infty} V(T_n) = \frac{\mu_4 - \mu_2^2}{4\mu_2}$$

$$\lim_{n \rightarrow \infty} \text{EQM}(T_n) = \frac{\mu_4 - \mu_2^2}{4\mu_2} + (\sqrt{\mu_2} - \sigma_0)^2$$

où μ_k est le moment centré sur la moyenne, d'ordre k.

$$\begin{aligned} \mu_4 = & (1-\epsilon) \int_{-\infty}^{+\infty} (x-c)^4 \varphi_0(x; m_0, \sigma_0) dx + \\ & + \epsilon \int_{-\infty}^{+\infty} (x-c)^4 \varphi_1(x; m_1, \sigma_1) dx \end{aligned}$$

$$\text{où } c = (1-\epsilon)m_0 + \epsilon m_1 = \mu_1'$$

Pour $\sigma_0 = 1$ et $m_0 = 0$, on obtient

$$\begin{aligned} \mu_4 = & (1-\epsilon)(c^4 + 6c^2 + 3) + \epsilon(3\sigma_1^4 + 6\sigma_1^2 m_1^2 + m_1^4 - 12c\sigma_1^2 m_1 - \\ & - 4c m_1^3 + 6c^2 \sigma_1^2 + 6c^2 m_1^2 - 4c^3 m_1 + c^4) \end{aligned}$$

De même, pour μ_2 , on obtient :

$$\mu_2 = (1-\epsilon)(c^2 + 1) + \epsilon(\sigma_1^2 + m_1^2 - 2c m_1 + c^2)$$

b) Modèle de contamination exponentiel

$$\mu_4 = (1-\varepsilon) \int_{-\infty}^{+\infty} (x-c)^4 \varphi_0(x; 0, 1) dx + \mu \int_a^{+\infty} (x-c)^4 e^{-\lambda(x-a)} dx - (1-\varepsilon) \int_a^{+\infty} (x-c)^4 \varphi_0(x; 0, 1) dx$$

$$\text{où } c = \frac{\mu}{\lambda^2} (\alpha\lambda + 1 - \lambda^2) = \mu_1'$$

Après développement, l'expression devient :

$$\begin{aligned} \mu_4 = & (1-\varepsilon)(c^4 + 6c^2 + 3) + \mu \left[\frac{(\alpha-c)^4}{\lambda} + \frac{4(\alpha-c)^3}{\lambda^2} + \right. \\ & \left. + \frac{12(\alpha-c)^2}{\lambda^3} + \frac{24(\alpha-c)}{\lambda^4} + \frac{24}{\lambda^5} \right] - \\ & - (1-\varepsilon) [\alpha^3 \varphi_0(\alpha) + 3(\alpha \varphi_0(\alpha) + 1 - \phi_0(\alpha))] - \\ & - 4c(\alpha^2 \varphi_0(\alpha) + 2\varphi_0(\alpha)) + 6c^2(\alpha \varphi_0(\alpha) + 1 - \phi_0(\alpha)) - \\ & - 4c^3 \varphi_0(\alpha) + c^4(1 - \phi_0(\alpha)) \end{aligned}$$

$$\begin{aligned} \mu_2 = & (1-\varepsilon)(c^2 + 1) + \mu \left[\frac{(\alpha-c)^2}{\lambda} + \frac{2(\alpha-c)}{\lambda^2} + \frac{2}{\lambda^3} \right] - \\ & - (1-\varepsilon) [\alpha \varphi_0(\alpha) + 1 - \phi_0(\alpha) - 2c \varphi_0(\alpha) + c^2(1 - \phi_0(\alpha))] \end{aligned}$$

$$\text{où } \phi_0(x) = \int_{-\infty}^x \varphi_0(u; 0, 1) du \quad \text{et } \varphi_\alpha(x) = \varphi_0(x; 0, 1)$$

Comme en 2.a), l'EQM asymptotique est égal à

$$\frac{\mu_4 - \mu_2^2}{4\mu_2} + (\sqrt{\mu_2} - 1)^2$$

3) Ecart absolu moyen

a) Modèle de contamination gaussien

$$\begin{aligned} \lim_{n \rightarrow \infty} E(T_n) &= \sqrt{\frac{\pi}{2}} \int_{-\infty}^{+\infty} |x-c| f(x) dx \\ &= \sqrt{\frac{\pi}{2}} \left[-(1-\varepsilon) \int_{-\infty}^c (x-c) \varphi_0(x) dx - \varepsilon \int_{-\infty}^c (x-c) \varphi_1(x) dx + \right. \\ &\quad \left. + (1-\varepsilon) \int_c^{+\infty} (x-c) \varphi_0(x) dx + \varepsilon \int_c^{+\infty} (x-c) \varphi_1(x) dx \right] \end{aligned}$$

où $\varphi_i(x) = \varphi_i(x; m_i, \sigma_i)$. Pour $m_0=0$ et $\sigma_0=1$, on obtient après développement :

$$\begin{aligned} \lim_{n \rightarrow \infty} E(T_n) &= \sqrt{\frac{\pi}{2}} \left[-(1-\varepsilon) (-2\varphi_0(c) - 2c\phi_0(c) + c) - \right. \\ &\quad \left. - \varepsilon (2m_1\phi_0(c') - m_1 - 2c\phi_0(c') - \sigma_1\varphi_0(c') + c) \right] \end{aligned}$$

$$\text{où } c' = \frac{c - m_1}{\sigma_1}$$

$$\lim_{n \rightarrow \infty} \text{biais}(T_n) = \lim_{n \rightarrow \infty} E(T_n) - 1$$

$$\lim_{n \rightarrow \infty} V(T_n) = \frac{\pi}{2} \mu_2 - \left(\lim_{n \rightarrow \infty} E(T_n) \right)^2$$

$$\lim_{n \rightarrow \infty} \text{EQM}(T_n) = \frac{\pi}{2} \mu_2 - 2 \lim_{n \rightarrow \infty} E(T_n) + 1$$

μ_2 a déjà été calculé en 2.a) .

b) Modèle de contamination exponentiel

$$\begin{aligned}
 \lim_{n \rightarrow \infty} E(T_n) = & \sqrt{\frac{\pi}{2}} \{ -(1-\epsilon) \int_{-\infty}^c (x-c) \varphi_0(x; 0, 1) dx + \\
 & + (1-\epsilon) \int_c^{+\infty} (x-c) \varphi_0(x; 0, 1) dx - \\
 & - \int_a^{\max(a, c)} (x-c) l e(x) dx + \\
 & + (1-\epsilon) \int_a^{\max(a, c)} (x-c) \varphi_0(x; 0, 1) dx + \\
 & + \int_{\max(a, c)}^{\infty} (x-c) l e(x) dx - \\
 & - (1-\epsilon) \int_{\max(a, c)}^{\infty} (x-c) \varphi_0(x; 0, 1) dx \}
 \end{aligned}$$

où $l e(x) = \mu e^{-\lambda(x-a)}$. Afin de faciliter les calculs, nous intraduisons la fonction indicatrice

$$I_a(x) = \begin{cases} 1 & \text{si } x > a \\ 0 & \text{si } x \leq a \end{cases}$$

Alors, après développement, l'expression ci-dessus devient :

$$\begin{aligned}
 & \sqrt{\frac{\pi}{2}} \{ (1-\epsilon) (\varphi_0(c) + c \phi_0(c) + \varphi_0(c) - c(1-\phi_0(c))) + \\
 & + I_a(c) \left[\frac{c}{\lambda} l e(c) + \frac{\mu \sigma}{\lambda} - \frac{1}{\lambda^2} l e(c) + \frac{\mu}{\lambda^2} + (1-\epsilon) (\varphi_0(a) - \varphi_0(c)) + \right. \\
 & + \left. \frac{c}{\lambda} (\mu - l e(c)) - (1-\epsilon) (\phi_0(c) - \phi_0(a)) \right] + \\
 & + (1 - I_a(c)) \left[\frac{\mu \sigma}{\lambda} + \frac{\mu}{\lambda^2} - (1-\epsilon) \varphi_0(a) - \frac{c \mu}{\lambda} + c(1-\epsilon) (1 - \phi_0(a)) \right]
 \end{aligned}$$

$$+ I_{\alpha}(c) \left[\frac{c}{\lambda} l e(c) + \frac{1}{\lambda^2} l e(c) - (1-\varepsilon) \varphi_0(c) - \right. \\ \left. - \frac{c}{\lambda} l e(c) + c(1-\varepsilon)(1-\varphi_0(c)) \right]$$

où $\phi_0(\cdot)$ et $\varphi_0(\cdot)$ sont définis en 2.b.

$$\lim_{n \rightarrow \infty} EQM(T_n) = \frac{\pi}{2} \mu_2 - 2 \lim_{n \rightarrow \infty} E(T_n) + 1 .$$

μ_2 a été calculé en 2.b) .

Références bibliographiques

- Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W. (1972). Robust estimates of Location. Princeton Univ. Press.
- Anscombe, F.J. (1960). Rejection of Outliers. Technometrics 2, 123-147.
- Anscombe, F.J., Tukey, J.W. (1963). The examination and analysis of residuals. Technometrics 5, 141-160.
- Collins, J.R. (1976). Robust estimation of a location parameter in the presence of asymmetry. Ann. Statist. 4, 68-85.
- Denby, L., Martin, R.D. (1979). Robust estimation of the first-order autoregressive parameter. Josa 74, 140-146.
- Dodge, Y. (1982). Robust estimation of regression coefficients by minimizing convex combination of L_1 and L_2 . In : Compstat 1982 2, 81-82.
- Fuller, W.A. (1976). Introduction to Statistical Time Series. J. Wiley.
- Hampel, F.R. (1971). A general qualitative definition of robustness. Ann. Math. Statist. 42, 1887-1896.
- Hampel, F.R. (1974). The influence curve and its role in robust estimation. Josa 69, 383-393.
- Huber, P.J. (1964). Robust estimation of a location parameter. Ann. Statist. 5, 73-101.
- Huber, P.J. (1967). The Behavior of maximum Likelihood estimates under nonstandard conditions. Fifth Berkeley Symposium on mathematical statistics and probability 1, 221-233.

- Huber, P.J. (1969). Théorie de l'inférence statistique robuste. Les Presses de l'Université de Montréal.
- Huber, P.J. (1973). Robust regression : asymptotics, conjectures and Monte Carlo. Ann. Statist. 1, 799-821.
- Huber, P.J. (1977). Robust Statistical procedures. Society for industrial and applied mathematics.
- Huber, P.J. (1981). Robust Statistics. J. Wiley.
- Jaeckel, L.A. (1971). Robust estimates of location : symmetry and asymmetric contamination. Ann. Math. Statist. 42, 1020-1034.
- Kendall, M.G., Stuart, A. (1969). The advanced theory of statistics 1. Charles Griffin and Co, London.
- Kendall, M.G., Stuart, A. (1973). The advanced theory of statistics 2. Charles Griffin and Co, London.
- Kendall, M.G., Stuart, A. (1976). The advanced theory of statistics 3. Charles Griffin and Co, London.
- Lejeune, M. (1982). Estimation de densité à noyau variable. Cahiers de Méthodes Quantitatives, Univ. de Neuchâtel.
- Martin, R.D., Jang, J. (1976). Asymptotic properties of robust generalized M- estimates for the first order autoregressive parameter. Techn. memor., Bell Laboratories, Murray Hill, N.J.
- Martin, R.D., Zeh, J.E. (1978). Generalized M- estimates for autoregressions, including small-sample efficiency robustness. Technical report 214, Dep. of electrical engineering, Univ. of Washington.

- Martin, R.D. (1981). The Cramer-Rao bound and robust M- estimates for autoregressions. Technical report 9, Dep. of Statistics, Univ. of Washington.
- Tukey, J.W. (1960). A survey of sampling from contaminated distributions. In : Contributions to Probability and Statistics. Stanford Univ. Press.