Enveloppement dans les modèles de régression paramétriques et non-paramétriques

Christophe Bontemps

20 Décembre 1994

Introduction

"As a research tactic, encompassing provides a basis for model comparisons, as well as integrating a large and diverse literature covering nested and non-nested hypothesis tests"

David F. Hendry et Jean-François Richard (1986)

Une des plus importantes activités scientifiques a été, et est toujours, la comparaison de théories et de modèles. Il est en effet extrêmement rare qu'un phénomène soit expliqué complètement par une théorie unique faisant l'unanimité. L'histoire des sciences connaît de nombreux exemples de luttes entre partisans de théories contradictoires, le temps seul parvenant à désigner les vainqueurs. De nos jours, si une théorie est acceptée comme utile et potentiellement durable, il est important de la confronter avec la réalité d'expériences, ou de données, ce qui est le rôle de la statistique. Toutefois une des faiblesses de cette discipline est qu'elle ne s'est intéressée que récemment à la validation de théories. Les études statistiques en économie, par exemple, mènent souvent à des situations conflictuelles, les conclusions s'opposant les unes aux autres, sans donner de méthode effective pour décider quelle théorie adopter. L'idée qu'une théorie nouvelle doit apporter un *progrès* dans la connaissance d'un phénomène est évidemment mise en avant, ce progrès est souvent jugé par sa capacité à expliquer des éléments que les autres théories, plus anciennes, n'expliquent pas. Toutefois, il semble stratégiquement important de s'assurer de la capacité d'une nouvelle théorie à expliquer également ce que les autres théories expliquaient déjà.

L'idée qu'une théorie doit être capable d'incorporer les résultats obtenus par des théories concurrentes, bien qu'adoptée implicitement par de nombreux scientifiques, n'a été formalisée que récemment en statistique sous le terme de "principe d'enveloppement"[1], au travers des travaux de Florens, Hendry, Mizon et Richard, d'une part (voir Mizon [65], Mizon et Richard [66] et Hendry et

[1] mot que nous choisissons pour la traduction de "*encompassing*"

Richard [54]), et de ceux de Gourieroux, Monfort et Trognon, d'autre part (voir Gourieroux et Monfort[38] et [39] ainsi que Gourieroux, Monfort et Trognon [42]) . L'extension de ces travaux au cadre bayésien, relié à la notion de spécificité (voir Florens, Hendry et Richard [31]), présente une vision unificatrice de cette notion, l'enveloppement bayésien présentant les mêmes caractéristiques que l'enveloppement classique. L'apport de Gourieroux, Monfort et Trognon [42] dans un contexte dynamique a permis l'introduction d'une procédure opérationnelle d'information indirecte [40]. L'ensemble de ces auteurs préconise également l'emploi de tests basés sur ce principe, et plus particulièrement Hendry [53].

L'étude de l'enveloppement est l'objet du premier chapitre, où nous discuterons des définitions exactes formalisant ce principe, toutefois une brève discussion informelle clarifie l'analyse.

Quel est le "vrai " modèle ?

Lorsque l'on parle de choix de modèles on est souvent amené à supposer qu'il existe un "vrai " modèle ayant engendré les données. Bien qu'inconnu et d'une complexité telle que sa connaissance exacte ne peut être envisagée, ce processus de génération des données fait l'objet d'hypothèses plus ou moins précises : il peut être spécifié paramétriquement ou non-paramétriquement, il peut appartenir à l'un des modèles ou être extérieur , il peut être dynamique ou pas, stationnaire ou non, etc... Conformément à Florens, Hendry et Richard [31], nous définirons séparément le "*processus de génération des données" et* les "*modèles".*

Le *processus de génération des données* est le mécanisme inconnu dont sont issues les observations, conceptuellement, c'est un élément $\mathcal{P}_0$ d'une classe de probabilités $\mathcal{P} = \{P_\theta, \theta \in \Theta_0\}$ sur l'espace mesurable $(\Omega, \mathcal{A})$. Θ_0 est l'espace paramétrique indexant $\mathcal{P}$, il peut éventuellement être fonctionnel et, tout comme $\mathcal{P}$, ne sera pas explicitement spécifié. $\mathcal{P}$ peut être défini de manière très large, par exemple comme l'ensemble des lois de probabilités admettant leurs 2 premiers moments.

Par "*modèle* $\mathcal{M}$ " nous entendrons le couple constitué d'un modèle d'estimation d'un paramètre d'intérêt, $\delta \in \Theta_\delta$, (Θ_δ étant typiquement de dimension inférieure à celle de Θ_0, pourra également être fonctionnel), et d'un estimateur. Il faudrait donc noter $(\mathcal{M}, \widehat{\delta})$, au lieu de $\mathcal{M}$, toutefois, après avoir levé toute ambiguïté , nous ignorerons cette notation.

On cherche à confronter un modèle $\left(\mathcal{M}_1, \widehat{\beta}\right)$ avec un modèle rival $(\mathcal{M}_2, \widehat{\gamma})$, où $\widehat{\beta}$ et $\widehat{\gamma}$ sont deux estimateurs convergents des paramètres β et γ respectivement, appartenant aux espaces paramétiques, ou fonctionnels Θ_β et Θ_γ ; ces deux espaces pouvant avoir des dimensions différentes.

Le modèle $\mathcal{M}_1$ enveloppe le modèle $\mathcal{M}_2$ s'il existe une "*fonction de lien*", Γ permettant de retrouver $\widehat{\gamma}$ à partir de $\widehat{\beta}$, c'est-à-dire, telle que l'on puisse

retrouver les résultats de $\mathcal{M}_2$ par ceux de $\mathcal{M}_1$.

Dans ce contexte d'enveloppement, l'approche de Gourieroux et Monfort [39] présente l'originalité de supposer le processus de génération des données extérieur aux modèles en présence. Cette étude propose ainsi le problème de choix entre deux modèles, deux approximations du vrai modèle, de manière symétrique, aucun des deux modèles n'ayant de rôle privilégié. L'enveloppement est alors envisagé dans un sens ($\mathcal{M}_1$ enveloppe $\mathcal{M}_2$) comme dans l'autre ($\mathcal{M}_2$ enveloppe $\mathcal{M}_1$), les deux sens n'étant pas forcément incompatibles.

Un autre point de vue est de considérer l'un des deux modèles comme un favori que l'on cherche à confronter avec un autre modèle, l'intérêt est alors la validation de ce modèle plutôt que du choix pur entre modèles concurrents[2]. Dans des situations pratiques, où les modèles sont inévitablement mal-spécifiés, il est souvent plus informatif d'analyser les forces et faiblesses respectives de chacun, que de chercher à sélectionner l'un des modèles. De même, le fait qu'un modèle $\mathcal{M}_1$ n'enveloppe pas un concurrent $\mathcal{M}_2$, indique que ce dernier incorpore des caractéristiques spécifiques qui n'ont pas été prises en compte par $\mathcal{M}_1$. Au lieu de rejeter simplement un tel modèle, cette faiblesse peut être exploitée plus constructivement, en incorporant les caractéristiques pertinentes relevées par $\mathcal{M}_2$ et ainsi améliorer la connaissance du phénomène étudié, c'est-à-dire progresser. Nous suivrons Hendry et Richard [54] dans cette voie, où l'enveloppement relève plus de la *comparaison* de modèles que du *choix* de modèles.

Enveloppement exact ou approché ?

L'enveloppement ("e*xact*"), tel que nous venons de le définir, n'est, en général, pas vérifié. Dans ce cas, il est toutefois possible de mesurer le défaut d'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$. Il nous faut pour cela introduire de manière plus précise la correspondance liant les résultats de $\mathcal{M}_1$ avec ceux de $\mathcal{M}_2$.

L'utilisation du critère d'information de Kulback-Leibler [57](KLIC), dans un contexte de maximum de vraisemblance, permet de définir une telle correspondance entre Θ_β et Θ_γ. Dans la lignée de Sawa [77], la pseudo-vraie valeur est définie comme l'élément (s'il existe) minimisant le KLIC. Cette définition, qui figure également chez White [90] ou Gourieroux, Monfort et Trognon [42], semble avoir été introduite (implicitement) dans l'oeuvre de Cox [21] et [22] relative aux tests d'hypothèses non-emboîtées, ainsi que dans les travaux de Huber [55].

La différence entre l'estimateur $\widehat{\gamma}$ et la pseudo-vraie valeur, ou un estimateur de celle-ci, permet une mesure du défaut d'enveloppement exact, et

[2]Cette vision directionnelle correspond à l'idée de confronter une théorie nouvelle à une théorie déjà éprouvée

définit l'enveloppement approché. Celui-ci sera réalisé lorsque cette différence, ou une fonction de cette différence, sera nulle, presque sûrement ou asymptotiquement.

De même, dans un contexte bayésien, l'enveloppement exact basé sur l'existence d'une correspondance entre les *a posteriori* des deux modélisateurs, ne sera que rarement vérifié. Un concept de spécificité sera défini afin de mesurer le défaut d'enveloppement qui s'exprimera comme une "*distance incompressible*" entre modèles. La pseudo-vraie valeur bayésienne sera elle aussi définie comme réalisant le minimum de la spécificité entre les modèles.

Dans la pratique (qu'elle soit classique ou bayésienne), l'enveloppement se jugera sur l'enveloppement approché. Ainsi les différents tests analysés dans le chapitre 2, seront basés sur la recherche de la nullité du défaut d'enveloppement exact, c'est à dire sur l'enveloppement approché. La littérature économétrique s'est d'ailleurs principalement concentrée sur cette définition plus opérationnelle.

Classique ou bayésien ?

Les modèles bayésiens se distinguent des modèles classiques en incorporant une densité à priori sur les paramètres, ce qui représente une extension des modèles classiques à un cadre où l'on dispose d'un ensemble d'information plus vaste. Le but de l'apprentissage bayésien est alors de passer de l'*a priori* sur le paramètre, à l'*a posteriori* (conditionnel à l'échantillon), par l'utilisation judicieuse de la règle de Bayes sur la loi jointe à l'échantillon et au paramètre. L'intérêt du modèle reposant sur cet *a posteriori,* il est alors naturel de baser la notion d'enveloppement, en tant que comparaison de modèles, sur l'étude des *a posteriori* de chacun des modèles.

Il est remarquable que la notion d'enveloppement s'étende aussi naturellement au cadre bayésien. En effet, la définition de l'enveloppement d'un modèle par un autre y est pratiquement la même, les estimateurs classiques proposés informellement ici seront remplacés par des densités *a posteriori*, la fonction de lien Γ devant être remplacée par une probabilité de transition.

En fait, dans un contexte probabiliste que nous ne détaillerons pas ici, le concept de probabilité de transition réunit les deux approches classique et bayésienne.

La principale difficulté de cette généralisation de l'enveloppement consiste en la recherche de la probabilité de transition donnant la pseudo-vraie valeur bayésienne (voir section 1.4). La complexité des calculs de celle-ci pose un réel problème d'estimation. Cette difficulté peut être contournée par l'utilisation de techniques de simulation, comme l'échantillonneur de Gibbs, (voir Bouoiyour [13]) , ou par des techniques d'approximation qui permettent un calcul opérationnel (voir Florens, Hendry et Richard [31]). Malheureusement, ces procédures ne sont encore définies que pour des cas particuliers (voir Florens, Larribeau et Mouchart [33]).

Asymptotique ou fini ?

La propriété d'enveloppement est essentiellement une propriété de "petit échantillon", typiquement cette notion trouve sa place naturelle dans un contexte bayésien c'est-à-dire appliqué à des échantillons finis. Cependant, l'approche asymptotique sera privilégiée dans ce travail. Tout d'abord, pour être opérationnelle, la propriété d'enveloppement doit pouvoir être testée. Ces tests qui ont été élaborés dans la littérature sur les problèmes de spécification sont majoritairement asymptotiques (voir Hausman [52] et White [91] entre autres). Il est donc nécessaire d'effectuer un minimum de théorie asymptotique afin de déterminer les lois des statistiques de test intervenant dans ce contexte. D'autre part, le calcul des pseudo-vraies valeurs est souvent simplifié asymptotiquement. Gouriéroux, Monfort et Trognon [42] proposent cependant des procédures de test basées sur des pseudo-vraies valeurs finies. Ces auteurs mettent en avant l'importance de ces pseudo-vraies valeurs finies dans des modèles conditionnels, et décrivent également les cas particuliers où celles-ci coïncident avec les pseudo-vraies valeurs asymptotiques. Dans l'optique du chapitre 4 où nous traiterons de modèles (et donc d'estimateurs) fonctionnels, l'approche asymptotique sera bien évidemment privilégiée.

Emboîtés ou non-emboîtés ?

Dans son article sur le problème général de la sélection de modèles, Pesaran [70] écrit : "*In many economic applications the models that we eventually encounter are often non-nested in the sense that they have separate parametric families and one model cannot be obtained from the others as a limiting process. Unfortunately, in such cases the application of the classical likelihood-ratio test procedure will not be correct and other suitable methods of testing have to be sought*". Des procédures ont ainsi été examinées par de nombreux auteurs, afin de réconcilier les modèles non-emboîtés avec les techniques existantes pour les modèles emboîtés. Cox ([21] et [22]), développe une procédure adaptée du test de rapport de vraisemblance. Cette méthode est basée sur l'examen, d'une part, des différences des log-vraisemblances empiriques, d'autre part la même différence est évaluée en supposant que $\mathcal{M}_1$ est "vrai" (voir Pesaran [70]).

Une des idées à été d'utiliser un "sur-modèle" emboîtant artificiellement les modèles concurrents. Cependant l'issue de ces procédures n'est pas satisfaisante puisque les deux modèles peuvent être simultanément acceptés ou rejetés, un autre problème est la forte collinéarité pouvant exister entre les variables explicatives intervenant dans le sur-modèle. Atkinson [4], reprend également l'idée d'un sur-modèle dont la densité est proportionnelle à une moyenne géométrique des densités des modèles concurrents. Davidson et Mac Kinnon [24], proposent un sur-modèle additif et contournent l'obstacle

de l'estimation séparée des paramètres des modèles et du paramètre liant les modèles (λ) en séquançant la procédure de test. On calcule d'abord les résidus issus de l'estimation de $\mathcal{M}_2$ que l'on reporte ensuite dans le sur-modèle où l'on peut alors tester de la nullité (ou l'égalité à 1) de λ, (voir section 2.1.3).

Hendry et Richard [54] notent que le principe d'enveloppement s'applique, que les modèles soient emboîtés ou non. Heuristiquement, un sur-modèle $\mathcal{M}_c$ emboîtant les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$, aura la même spécificité que $\mathcal{M}_2$ vis-à-vis du modèle $\mathcal{M}_1$ et ne saurait donc apporter aucune aide à la décision . Nous observerons sur un exemple, (exemple 3, section 1.3.1), la situation où $\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ est équivalent à $\mathcal{M}_1$ enveloppe $\mathcal{M}_c$. Dès lors, l'enveloppement parcimonieux, (voir section 1.3.1), permet d'envisager une procédure de réduction des modèles, l'objectif étant de construire des modèles "*plus simples*" qui présentent la même capacité à envelopper des modèles "*plus grands*" .

Ce travail se veut une contribution aux recherches en cours sur la notion d'enveloppement dans les modèles de régression. Les comportements asymptotiques des statistiques mesurant le défaut d'enveloppement sont maintenant bien connus dans le cadre paramétrique,et seront rappelés dans le chapitre 2. Notre objectif est d'étendre ces résultats au cadre de la régression non-paramétrique.

Les techniques d'estimation fonctionnelle de la régression, proposées chapitre 3, nous permettent en effet, d'envisager une extension de ces travaux à des modèles autres que linéaires et/ou gaussiens. Dans cette optique la question centrale que nous aborderons dans ce travail sera :

"Existe t'il des procédures de test d'enveloppement entre modèles de régression libres de toute forme fonctionnelle ?"

Cette question en appelle d'autres auxquelles nous tenterons de répondre, dans le chapitre 4, notamment :

Comment se comporte l'estimateur non-paramétrique d'un modèle de régression $\mathcal{M}_2$ sous l'hypothèse que $\mathcal{M}_1$ est "vrai" ?

Quelle statistique de test globale peut-on envisager pour tester de l'enveloppement dans ce cadre ?

Quelle en est la perte en terme de vitesse de convergence par rapport au cas paramétrique ?

Nous nous efforcerons de répondre à ces questions par les procédures développées dans le quatrième chapitre.

Nous chercherons également à comparer par enveloppement procédures paramétriques et non-paramétriques. Nous étudierons 4 cas en combinant les spécifications paramétriques et fonctionnelles pour chacun des deux modèles

en présence. Cette étude nous poussera à étudier de manière précise les choix arbitraires qui peuvent être faits dans la sélection des estimateurs de chacun des modèles. Ces choix, et particulièrement ceux des fenêtres, peuvent influer sur les critères nécessairement objectifs de comparaison de modèles, et seront mis en évidence. Les simulations conduites et proposées dans le chapitre 5 viendront étayer nos résultats.

Enfin et surtout, nous proposerons un critère global d'enveloppement dont la distribution asymptotique sera caractérisée. Ce critère convergera vers ce que nous appellerons "*une loi normale fuyante*", c'est-à-dire qu'un terme résiduel croissant s'ajoutera au terme donnant la normalité asymptotique dans notre critère. Cette caractéristique, propre au cadre non-paramétrique, nous indique que notre approche asymptotique comporte des faiblesses. Ces faiblesses pourraient être compensées dans le futur par l'utilisation de techniques de Bootstrap.

Chapter 1

Le principe d'enveloppement

"One model is said to encompass another if the former can account for, or explain, the results of the latter."

David F. Hendry et Jean-François Richard (1986)

1.1 Définition de l'enveloppement exact

Soit Y une variable aléatoire définie sur l'espace mesurable $(\Omega, \mathcal{A})$, et $Y_n = (y_i)_{i=1,\dots,n}$ n réalisations indépendantes de cette variable.

On cherche à confronter un modèle $\mathcal{M}_1$ candidat à la modélisation du processus de génération de données ou tout du moins candidat à la représentation d'aspects pertinents de ce processus, avec un modèle rival $\mathcal{M}_2$. Les deux modèles, indexés par les paramètres β et γ respectivement, reposent sur des espaces paramétriques, Θ_β et Θ_γ, qui peuvent éventuellement être fonctionnels.

Soit $\widehat{\beta}_n$ et $\widehat{\gamma}_n$ des estimateurs consistants de β et γ dans leurs modèles respectifs, les estimateurs $\widehat{\beta}_n$ et $\widehat{\gamma}_n$ dépendent de l'échantillon Y_n.

$\mathcal{M}_1$ étant le candidat que l'on cherche à confronter à $\mathcal{M}_2$, on va chercher à analyser sa capacité à "expliquer" $\mathcal{M}_2$, ou plutôt, sa capacité à expliquer les résultats de $\mathcal{M}_2$ par ses propres résultats. Pour cela nous proposons la définition suivante, donnée initialement par Hendry et Richard [54] :

Définition 1.1 *(Enveloppement exact) :*

On dira que "$\mathcal{M}_1$ enveloppe exactement $\mathcal{M}_2$ " ($\mathcal{M}_1 \mathcal{E}_e \mathcal{M}_2$) s'il existe Γ, "fonction de lien", $\Gamma : \Theta_\beta \longrightarrow \Theta_\gamma$, telle que, pour tout échantillon Y_n :

$$\widehat{\gamma}(Y_n) = \Gamma\left(\widehat{\beta}(Y_n)\right) \qquad (M_1 \quad p.s.) \tag{1.1}$$

Ceci signifie bien que l'on peut obtenir, à partir de l'estimation des paramètres de $\mathcal{M}_1$, les mêmes résultats que ceux obtenus par $\mathcal{M}_2$ puisqu'on obtient $\widehat{\gamma}(Y_n)$ à partir de $\widehat{\beta}(Y_n)$. $\mathcal{M}_1$ est donc préférable à $\mathcal{M}_2$ puisqu'il contient potentiellement les résultats de son concurrent.

Exemple 1

Soient les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ paramétrés par β et γ sur $\Re^+$ et représentés par les densités suivantes:

$$\mathcal{M}_1 : Y \sim \mathcal{N}(\beta, 1) \quad et \quad \mathcal{M}_2 : Y \sim \mathcal{N}(e^{\gamma}, 1)$$

munis des estimateurs

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad et \quad e^{\widehat{\gamma}} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Sur cet exemple $\mathcal{M}_2$ est une reparamétrisation de $\mathcal{M}_1$, et donc $\mathcal{M}_1$ *enveloppe exactement* $\mathcal{M}_2$, en effet la fonction $\Gamma(\cdot) = \log(\cdot)$ nous donne donc explicitement $\widehat{\gamma} = \Gamma(\widehat{\beta})$.

Il est à noter que l'on a ici une fonction Γ bijective sur $\Re^+$ et donc nous avons également $\widehat{\beta} = e^{\widehat{\gamma}}$ ce qui signifie également que $\mathcal{M}_2$ *enveloppe* $\mathcal{M}_1$, les deux sens n'étant pas incompatibles. □

Exemple 2

Soit $Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \end{pmatrix}$ un vecteur aléatoire sur $(\Re^2, \mathcal{B}_{\Re^2}, \lambda_2)$ et $Y_n = (y_1, y_2, \ldots, y_n)$, n réalisations indépendantes de cette variable.

Considérons les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$, définis par les densités normales suivantes :

$$\mathcal{M}_1 : Y \sim \mathcal{N}_2\left(\begin{pmatrix} \mu \\ \nu \end{pmatrix}, \sum\right) \quad et \quad \mathcal{M}_2 : Y \sim \mathcal{N}_2\left(\begin{pmatrix} \eta \\ 1 \end{pmatrix}, \sum\right)$$

où $\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$, matrice de variance-covariance, est connue.

Le paramètre $\beta = \begin{pmatrix} \mu \\ \nu \end{pmatrix}$ est estimé naturellement par $\widehat{\beta}$:

$$\widehat{\beta} = \begin{pmatrix} \widehat{\mu} \\ \widehat{\nu} \end{pmatrix} = \begin{pmatrix} \overline{y}^{(1)} \\ \overline{y}^{(2)} \end{pmatrix}$$

où $\overline{y}^1$ et $\overline{y}^2$ sont les moyennes empiriques:

$$\overline{y}^1 = \frac{1}{n}\sum_{i=1}^{n} y_i^{(1)} \quad et \quad \overline{y}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^{(2)}$$

Un estimateur de $\gamma = \begin{pmatrix} \eta \\ 1 \end{pmatrix}$ est $\widehat{\gamma}$ avec:

$$\widehat{\gamma} = \begin{pmatrix} \widehat{\eta} \\ 1 \end{pmatrix} = \begin{pmatrix} \overline{y}^{(1)} \\ 1 \end{pmatrix}$$

Nous pouvons donc clairement calculer $\widehat{\gamma}$ à partir de $\widehat{\beta}$, puisque $\widehat{\gamma} = \Gamma(\widehat{\beta})$ où Γ est la fonction:

$$\begin{array}{rccc} & \Re^2 & \longrightarrow & \Re^2 \\ \Gamma : & \begin{pmatrix} u \\ v \end{pmatrix} & \longrightarrow & \begin{pmatrix} u \\ 1 \end{pmatrix} \end{array}$$

Sur cet exemple trivial, nous voyons comment un sous-modèle $\mathcal{M}_2$ est enveloppé exactement par un modèle dont il est la restriction, la fonction Γ étant la représentation de la restriction sur l'espace des paramètres. Nous verrons par la suite, section 1.3, que des sous-modèles peuvent envelopper les modèles dont ils sont issus, ce qui, au regard du principe de parcimonie, présente un intérêt beaucoup plus grand.

Remarque:

- Dans l'exemple précédent on peut proposer sur $\mathcal{M}_2$ un autre estimateur de γ en prenant $\widehat{\widehat{\gamma}} \neq \widehat{\gamma}$. Il n'est pas évident que l'enveloppement soit également réalisé avec ce nouvel estimateur de γ, puisqu'en changeant d'estimateur, nous changeons le modèle $\mathcal{M}_2$.

- Il se peut également que l'enveloppement soit réalisé mais par une fonction de lien différente. Prenons par exemple $\widetilde{\gamma}$ estimateur du maximum de vraisemblance :

  $$\widetilde{\gamma} = \begin{pmatrix} \widetilde{\eta} \\ 1 \end{pmatrix} \quad avec \quad \widetilde{\eta} = \overline{y}^1 + \pi\left(\overline{y}^2 - 1\right)$$

  où l'expression de π est $\pi = \frac{\sigma_{12}}{\sigma_{22}}$.

  Nous pouvons encore calculer $\widetilde{\gamma}$ à partir de $\widehat{\beta}$, mais à l'aide de la fonction $\Gamma' \neq \Gamma$ où Γ' est la fonction:

  $$\begin{array}{rccc} & \Re^2 & \longrightarrow & \Re^2 \\ \Gamma' : & \begin{pmatrix} u \\ v \end{pmatrix} & \longrightarrow & \begin{pmatrix} u + \pi(v-1) \\ 1 \end{pmatrix} \end{array}$$

Nous avons, ici également, $\tilde{\gamma} = \Gamma'(\widehat{\beta})$, l'enveloppement est donc vérifié pour ce nouveau modèle avec ce nouvel estimateur mais nous avons changé de fonction de lien.

Sur cet exemple, nous remarquons donc que $(\mathcal{M}_1, \widehat{\beta})$ enveloppe le modèle $(\mathcal{M}_2, \widehat{\gamma})$ ainsi que le modèle $(\mathcal{M}_2, \tilde{\gamma})$. □

1.1.1 Version dynamique

Soient $\mathcal{M}_1$ et $\mathcal{M}_2$ deux modèles paramétriques dynamiques sans exogènes candidats à la modélisation de la densité d'un vecteur aléatoire Y_t. Les densités respectives de $\mathcal{M}_1$ et $\mathcal{M}_2$ sont :

$$f(y_t \mid Y_{t-1}, \beta) \quad et \quad g(y_t \mid Y_{t-1}, \gamma)$$

où β et γ appartiennent aux espaces paramétriques Θ_β et Θ_γ, et où la matrice Y_{t-1}, regroupe les observations "passées" : $Y_{t-1} = (y_{t-1}, y_{t-2}, \cdots, y_1)$.

On associe au modèle $\mathcal{M}_1$ l'estimateur $\widehat{\beta}_T$ de β basé sur l'échantillon de taille T, Y_T, de même $\widehat{\gamma}_T$ est l'estimateur de γ associé à $\mathcal{M}_2$.

Govaerts, Hendry et Richard [43], proposent la définition de l'enveloppement dynamique, dans le même esprit que la définition 1.1 :

Définition 1.2 *: "Le modèle dynamique $\mathcal{M}_1$ enveloppe exactement $\mathcal{M}_2$", s'il existe une séquence de fonctions Γ_T telle que :*

$$\widehat{\gamma}_T = \Gamma_T(\widehat{\beta}_T) \quad (\mathcal{M}_1 \quad p.s.)$$

Ici encore, et pour tout T, la connaissance de $\widehat{\beta}_T$ associée à celle des fonctions de lien Γ_T, permet la connaissance de l'estimateur de $\mathcal{M}_2$, $\widehat{\gamma}_T$. Le modèle $\mathcal{M}_1$ sera donc préféré, contenant, implicitement l'ensemble des résultats de son rival.

Cette définition ne diffère de la définition (statique) donnée en (1.1) que par l'aspect séquentiel que doit revêtir ici la fonction de lien Γ, remplacée ici par une succession de fonctions de liens.

1.1.2 Propriétés

Nous pouvons reformuler la définition 1.1 d'une manière plus visuelle en examinant les relations entre les espaces Ω, Θ_β et Θ_γ:

Définition 1.1 (bis) :

$\mathcal{M}_1$ enveloppe exactement $\mathcal{M}_2$ s'il existe une fonction Γ telle que le schéma formel représenté par la figure1.1 soit "fermé ".

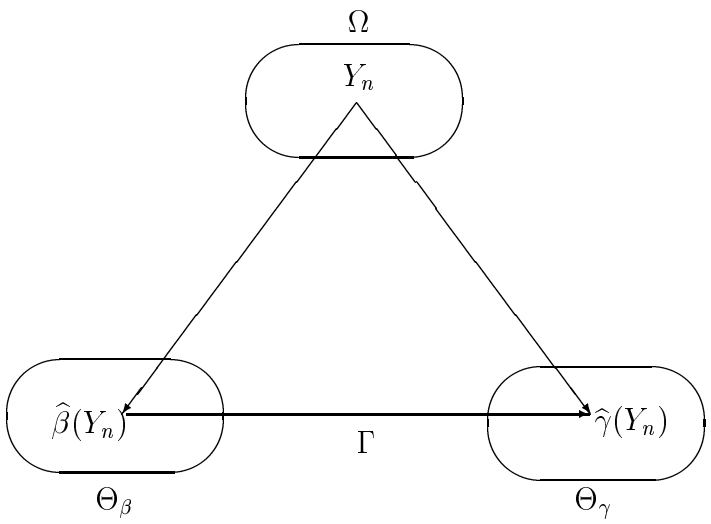

Figure 1.1: Enveloppement exact

Les espaces Θ_β et Θ_γ sur lesquels reposent les estimateurs $\widehat{\beta}_n$ et $\widehat{\gamma}_n$ issus de l'échantillon Y_n sont ainsi liés par la fonction Γ définissant la pseudo-vraie valeur $\Gamma(\widehat{\beta})$. Dès lors, le modèle 2 n'apporte rien que ne puissent expliquer les résultats du modèle 1.

Nous verrons section 1.4 que cette définition s'étend au cadre bayésien sans difficultés.

Nous obtenons quelques propriétés immédiates de cette définition :

- La relation d'enveloppement exact définie par (1.1) est transitive (voir ci-dessous)
- L'enveloppement exact, tel que nous le définissons ici, est une relation entre modèles estimés et non entre les modèles théoriques eux-mêmes.
- Cette relation ne dépend pas de la bonne ou de la mauvaise spécification des modèles en présence, chacun d'eux étant potentiellement mal-spécifié
- Si la fonction Γ est bijective alors l'enveloppement sera réciproque ($\mathcal{M}_1 \, \mathcal{E}_e \, \mathcal{M}_2$ et $\mathcal{M}_2 \; \mathcal{E}_e \, \mathcal{M}_1$), toutefois l'intérêt de comparer des modèles dont les paramètres sont en bijection est très limité (voir l'exemple 1).

Transitivité de l'enveloppement exact

La propriété d'enveloppement exact est une propriété transitive. Si un modèle $(\mathcal{M}_1, \widehat{\beta})$ enveloppe exactement un modèle $(\mathcal{M}_2, \widehat{\gamma})$, et si ce dernier enveloppe à son tour un modèle $(\mathcal{M}_3, \widehat{\delta})$, alors $\mathcal{M}_1$ enveloppe $\mathcal{M}_3$.

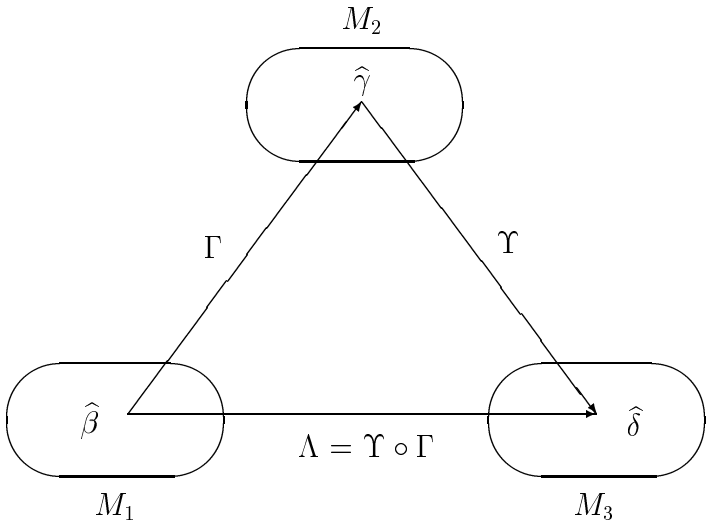

Figure 1.2: Transitivité de l'enveloppement exact

En effet, s'il existe Γ liant les espaces Θ_β et Θ_γ telle que $\widehat{\gamma} = \Gamma(\widehat{\beta})$

$$\Gamma : \quad \begin{array}{ccc} \Theta_\beta & \longrightarrow & \Theta_\gamma \\ \widehat{\beta} & \longrightarrow & \Gamma(\widehat{\beta}) = \widehat{\gamma} \end{array}$$

et s'il existe Υ liant les espaces Θ_γ et Θ_δ telle que $\widehat{\delta} = \Upsilon(\widehat{\gamma})$

$$\Upsilon : \quad \begin{array}{ccc} \Theta_\gamma & \longrightarrow & \Theta_\delta \\ \widehat{\gamma} & \longrightarrow & \Upsilon(\widehat{\gamma}) = \widehat{\delta} \end{array}$$

alors il existe $\Lambda = \Upsilon \circ \Gamma$ liant les espaces Θ_β et Θ_δ et telle que $\widehat{\delta} = \Lambda(\widehat{\beta})$

$$\Lambda = \Upsilon \circ \Gamma : \quad \begin{array}{ccc} \Theta_\beta & \longrightarrow & \Theta_\delta \\ \widehat{\beta} & \longrightarrow & \Lambda(\widehat{\beta}) = \widehat{\delta} \end{array}$$

Nous retrouvons la transitivité intuitive de cette notion. Plus visuellement nous avons le schéma (*fermé*) donné par la figure 1.2.

Il importe cependant d'être prudent : la définition de l'enveloppement fait intervenir des égalités *presque sûres*, pour des lois différentes.

En effet d'un côté on a :

$$\widehat{\gamma} = \Gamma(\widehat{\beta}) \quad \mathcal{M}_1 \ \textit{presque sûrement.}$$

et donc : $\left\{ Y_n \ \textit{tels que} \quad \widehat{\gamma}(Y_n) \neq \Gamma(\widehat{\beta}(Y_n)) \right\}$ est de mesure nulle pour $\mathcal{M}_1$.

D'autre part :

$$\widehat{\delta} = \Upsilon(\widehat{\gamma}) \quad \mathcal{M}_2 \ \textit{presque sûrement.}$$

c'est-à-dire que l'ensemble : $\left\{Y_n \; tels \; que \quad \widehat{\delta}(Y_n) \neq \Upsilon(\widehat{\gamma}(Y_n))\right\}$ est de mesure nulle pour $\mathcal{M}_2$.
L'égalité : $\widehat{\delta} = \Lambda(\widehat{\beta}) = \Upsilon \circ \Gamma(\widehat{\beta}) \quad \mathcal{M}_1 \; presque \; sûrement$
ne sera vérifiée que si l'ensemble $\left\{Y_n \; tels \; que \quad \widehat{\delta}(Y_n) \neq \Upsilon(\widehat{\gamma}(Y_n))\right\}$ est également de mesure nulle pour $\mathcal{M}_1$.

Dans un contexte paramétrique, sur des espaces réels par exemple, où les modèles sont définis par des lois de probabilités, il faut être prudent et imposer que $\mathcal{M}_2$ domine $\mathcal{M}_1$. Cette situation peut ne pas être réalisée pour des modèles de dimensions différentes ; typiquement si $\mathcal{M}_2$ est emboîté dans $\mathcal{M}_1$ et est de dimension inférieure, $\mathcal{M}_2$ ne dominera pas $\mathcal{M}_1$. Dans un cadre fonctionnel, il faudrait de même imposer aux négligeables de $\mathcal{M}_2$ de l'être pour $\mathcal{M}_1$ également.

Remarque :

Ces propriétés ne nous assurent pas de la pertinence du modèle enveloppant, en terme de modélisation du "vrai" processus de génération. De plus, le processus ayant engendré les données n'a pas la propriété d'envelopper toute tentative de modélisation basée sur Y_n. La notion d'enveloppement approché permet de récupérer cette propriété intuitive, la propriété de transitivité n'est, elle, pas conservée.

1.2 Enveloppement approché

D'une manière générale, l'enveloppement exact défini en (1.1) n'est vérifié que rarement en échantillon fini, et ce même si $\mathcal{M}_1$ est le processus de génération des données.

Face à ce constat, deux approches peuvent être envisagées, la première est basée sur une mesure du défaut d'enveloppement. Pour cela la pseudo-vraie valeur sera définie et reliée à la notion de spécificité entre modèles, typiquement la définition de l'enveloppement approché ne différera de l'enveloppement exact "que" par la détermination préalable de la fonction Γ.

La deuxième approche consiste à définir l'enveloppement asymptotiquement, la pseudo-vraie valeur étant définie comme une réinterprétation des paramètres de $\mathcal{M}_2$ sous l'éclairage de $\mathcal{M}_1$. Ces deux approches, bien que différenciées ici ne sont que deux visions approchées d'une notion exacte.

1.2.1 Principe général

Nous allons définir une "mesure" du défaut d'enveloppement qui servira de base à l'enveloppement approché. Le principe consiste à choisir une fonction réelle $\Psi(\widehat{\gamma}, \widehat{\beta})$ mesurant l'écart, ou la divergence, entre les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$.

La fonction de lien Γ

$$\begin{array}{rccc} \Gamma & : & \Theta_\beta & \longrightarrow & \Theta_\gamma \\ & & \widehat{\beta}(Y_n) & \longrightarrow & \Gamma(\widehat{\beta}(Y_n)) \end{array}$$

qui détermine la pseudo-vraie valeur $\Gamma(\widehat{\beta}(Y_n))$, est alors définie comme l'élément d'une classe de fonctions C_Φ, qui rapproche au mieux les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ au sens de cette mesure.

$$\Gamma(\widehat{\beta}) = \arg \min_{\Phi \in C_\Phi} \Psi(\widehat{\gamma}, \Phi(\widehat{\beta}))$$

Cette fonction Γ minimise la "spécificité" de $\mathcal{M}_2$ vis à vis de $\mathcal{M}_1$, c'est également celle qui donne le plus de possibilités à $\mathcal{M}_1$ d'expliquer les résultats de $\mathcal{M}_2$[1].

Il est essentiel de remarquer que selon les types de modèles examinés, selon les espaces "*paramétriques*", (qui peuvent être fonctionnels), selon les propriétés de C_Φ et selon les procédures d'estimation, la fonction de lien Γ (et donc la pseudo-vraie valeur) connaîtra des caractéristiques et des propriétés différentes.

Le principe général proposé ici peut être résumé par le programme suivant :

- Choix des estimateurs (c'est-à-dire choix des modèles soumis à l'étude)
- Définition de la fonction Ψ introduisant la spécificité
- Détermination et estimation de la pseudo-vraie valeur
- Calcul de la différence d'enveloppement minimale
- Test de la nullité de la spécificité (voir chapitre 2)

Dans les modèles paramétriques de maximum de vraisemblance, le critère d'information de Kulback-Leibler [57] est généralement adopté comme "*distance*" entre modèles [2]. Nous vérifierons que ce critère coïncide avec une mesure de la spécificité introduite par Florens, Hendry et Richard [31].

[1] La "spécificité" du modèle $\mathcal{M}_2$ vis-à-vis du modèle $\mathcal{M}_1$ est, en fait, la valeur minimale de la fonction Ψ, c'est-à-dire :

$$\Psi(\widehat{\gamma}, \Gamma(\widehat{\beta}))$$

Cette spécificité est évidemment conditionnelle à l'échantilon Y_n. Les tests d'enveloppement développés par la suite seront ensuite basés sur l'étude de la nullité de cette spécificité conditionnelle à l'échantillon.

[2] Ce critère porte souvent le nom de "*contraste*", exprimant ainsi l'idée qu'il s'agit d'un

Le contraste de Kulback et Leibler (KLIC)

Soit $(\Re, \mathcal{B}_\Re, \lambda)$ l'espace réel mesuré et Y une variable aléatoire réelle . Afin de nous assurer de l'existence de ce critère et pour obtenir des propriétés de régularité usuelles, nous devons introduire quelques notations et hypothèses sur les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$.

i) Le modèle $\mathcal{M}_1$ suppose que la variable Y admet la densité $f(y,\beta)$ par rapport à la mesure de Lebesgue λ, où f est continue en $\beta \in \Theta_\beta \subset \Re^l$

ii) Le modèle $\mathcal{M}_1$ suppose que Y admet la densité $g(y,\gamma)$ par rapport à λ, avec g continue en $\gamma \in \Theta_\gamma \subset \Re^m$

iii) le support de $f(y,\beta)$ est inclus dans celui de $g(y,\gamma)$.

Une mesure directionnelle de la distance entre $\mathcal{M}_1$ et $\mathcal{M}_2$ est donnée par le KLIC (Kulback-Leibler Information Criterion):

$$I(\mathcal{M}_1, \mathcal{M}_2) = \mathbf{E}_\beta \left[\log \left(\frac{f(y,\beta)}{g(y,\gamma)} \right) \right]$$

où $\mathbf{E}_\beta(\cdot)$ est l'espérance relative au modèle 1, i.e. à la densité $f(y,\beta)$.

Il est bon de remarquer que :

- Ce critère n'est pas une distance, puisque l'inégalité triangulaire n'est pas vérifiée et que $I(\mathcal{M}_1, \mathcal{M}_2) \neq I(\mathcal{M}_2, \mathcal{M}_1)$ en général.
- Ce critère est positif
- $I(\mathcal{M}_1, \mathcal{M}_2) = 0 \Leftrightarrow f(y,\beta) = g(y,\gamma)$

Preuve : Nous empruntons ce résultat à Gourieroux et Monfort [37].

L'inégalité de Jensen appliquée à la fonction convexe $-\log(x)$ nous donne :

$$\mathbf{E}_\beta \left[\log \left(\frac{f(y,\beta)}{g(y,\gamma)} \right) \right] = -\mathbf{E}_\beta \left[\log \left(\frac{g(y,\gamma)}{f(y,\beta)} \right) \right] \geq -\log \mathbf{E}_\beta \left(\frac{g(y,\gamma)}{f(y,\beta)} \right)$$

Or

$$-\log \mathbf{E}_\beta \left(\frac{g(y,\gamma)}{f(y,\beta)} \right) = -\log \int \frac{g(y,\gamma)}{f(y,\beta)} \cdot f(y,\beta)\, dy = 0$$

éclairage particulier (celui de $\mathcal{M}_1$), sur le rapport des vraisemblances. Le terme de "*divergence*" est également employé pour affirmer la notion d'écart entre modèles.

De plus, la fonction $-\log(x)$ étant strictement convexe, l'égalité à zéro n'a lieu que si $g(y,\gamma)/f(y,\beta)$ est égal à une constante, k. Comme $\mathbf{E}_\beta\left(\frac{g(y,\gamma)}{f(y,\beta)}\right) = 1$, on en déduit que $k = 1$. □

Les modèles de maximum de vraisemblance constituent un bon exemple de mise en oeuvre du principe général que nous adoptons pour définir l'enveloppement approché, nous resterons dans ce cadre tout au long de cette section. Dans ce contexte, Florens et alii [31] proposent de définir la pseudo-vraie valeur comme minimisant la *spécificité* de $(\mathcal{M}_2, \widehat{\gamma})$ vis-à-vis de $(\mathcal{M}_1, \widehat{\beta})$.

Une mesure de *la spécificité* de $(\mathcal{M}_2, \widehat{\gamma})$ par rapport à $(\mathcal{M}_1, \widehat{\beta})$ est donnée pour une fonction Γ par :

$$D_\Gamma(Y_n) = \int_\Omega \log\left[\frac{g(y, \widehat{\gamma}(Y_n))}{g(y, \Gamma(\widehat{\beta}(Y_n)))}\right] \; f(y,\beta) \; \lambda(dy)$$

Cette mesure est évidemment dépendante de l'échantillon Y_n[3].

La pseudo-vraie valeur $\Gamma(\beta)$ est définie comme réalisant le minimum du critère $D_\Gamma(Y_n)$ pour tout Y_n.

$$\Gamma(\beta) = \arg\min_\delta \int_\Omega \log\left[\frac{g(y,\widehat{\gamma})}{g(y,\delta)}\right] \; f(y,\beta) \; \lambda(dy)$$

Il est important de noter que, par cette définition de la pseudo-vraie valeur, nous cherchons volontairement à réduire au maximum la spécificité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$. En minimisant cette spécificité nous offrons ainsi la "plus faible résistance possible" à l'enveloppement de $\mathcal{M}_2$.

Il est aisé de voir que $\Gamma(\beta)$ réalise le minimum du critère d'information de Kulback-Leibler (KLIC).

Preuve :

Notons que, comme $\widehat{\gamma}$ est l'estimateur du maximum de vraisemblance, on a :

$$\frac{g(y,\widehat{\gamma})}{g(y,\delta)} \geq 1 \qquad \forall \delta \in \Theta_\delta$$

[3]Une manière de s'affranchir de l'échantillon consiste à introduire une probabilité sur l'espace (Ω,$\mathcal{A}$) dont le choix dépendra du cadre de travail (classique ou bayésien, paramétrique ou non-paramétrique, etc..)

Ensuite, par un simple jeu d'écriture, on obtient :

$$\begin{aligned}\Gamma(\beta) &= \arg\min_\delta \int \log\left[\frac{g(y,\widehat{\gamma})}{g(y,\delta)}\right] \; f(y,\beta)\, dy \\ &= \arg\max_\delta \int \log\left[g(y,\delta)\right] \; f(y,\beta)\, dy \\ &= \arg\min_\delta \; \int \log\left[\frac{f(y,\beta)}{g(y,\delta)}\right] \; f(y,\beta)\, dy\end{aligned}$$

Le dernier terme est bien le contraste de Kulback-Leibler. □

Nous vérifions ainsi que, dans le cadre présent de maximum de vraisemblance, minimiser la spécificité d'un modèle vis-à-vis d'un autre, revient à minimiser la distance qui les sépare au sens de Kulback-Leibler. Des mesures de spécificité autres que celle proposée ici peuvent être introduites, elles conduisent évidemment à d'autres pseudo-vraies valeurs et à d'autres tests.

Dans un contexte bayésien, on aura le souci de définir une "*spécificité inconditionnelle* " en supprimant la dépendance vis-à-vis de l'échantillon par intégration en y suivant la loi supposée de y (voir section 1.4). La mesure précédente est cependant préférée, elle conduit en effet à une présentation naturelle des distances entre modèles.

1.2.2 Définition de la pseudo-vraie valeur

Gourieroux, Monfort et Trognon [42], sur les bases des travaux de Sawa [77], Huber [55] et Cox ([21] et [22]), proposent en 83, la définition de la pseudo-vraie valeur dans le contexte présent de maximum de vraisemblance par :

$$\Gamma(\beta) = \arg\min_\delta \int_\Omega \log\left[\frac{f(y,\beta)}{g(y,\delta)}\right] f(y,\beta)\lambda(dy)$$

C'est-à-dire que la pseudo-vraie valeur associée à une procédure de maximum de vraisemblance est définie comme la valeur minimisant le KLIC $I(\mathcal{M}_1, \mathcal{M}_2)$ ce qui équivaut à minimiser la spécificité introduite ci-dessus.

Une autre expression équivalente est :

$$\Gamma(\beta) = \arg\max_\delta \int_\Omega \log\left[g(y,\delta)\right] f(y,\beta)\lambda(dy) \tag{1.2}$$

La fonction de lien Γ n'est donc pas définie analytiquement, mais résulte d'une procédure de minimisation. Sawa s'est le premier intéressé au calcul des pseudo-vraies valeurs, il montre (lemme 3.2), que la pseudo-vraie valeur $\Gamma(\beta)$ s'écrit également :

$$\Gamma(\beta) = E_\beta(\widehat{\gamma}) \tag{1.3}$$

où $E_\beta(\cdot)$ désigne l'espérance relative au modèle $\mathcal{M}_1$.

Si $\widehat{\gamma}$ est l'estimateur du maximum de vraisemblance du modèle paramétrique $\mathcal{M}_2$, on a :

$$\Gamma(\beta) = \int_\Omega \arg\max_\delta \log\left[g(y,\delta)\right] f(y,\beta)\lambda(dy) \tag{1.4}$$

La distinction entre (1.2) et (1.4), réside alors dans l'ordre des opérateurs.

Conformément à Hendry, Mizon et Richard (Voir Mizon [65], Mizon et Richard [66], ou Hendry et Richard [54]), l'espérance sous $\mathcal{M}_1$ de $\widehat{\gamma}$ définissant la pseudo-vraie valeur dans l'expression 1.3 est remplacée par :

$$\Gamma(\beta) = p\lim_{\mathcal{M}_1} \widehat{\gamma}$$

$\Gamma(\beta)$ se présente ici comme une réinterprétation de l'estimateur $\widehat{\gamma}$ par $\mathcal{M}_1$, elle est aisément estimable, dès lors que β l'est, par $\Gamma(\widehat{\beta})$. Nous utiliserons cette expression de la pseudo-vraie valeur dans la suite de ce travail.

Gourieroux et alii [42], proposent également une définition de la pseudo-vraie valeur en échantillon fini dont $\Gamma(\beta)$ est la limite.

1.2.3 Pseudo-vraie valeur à distance finie

Considérons l'échantillon constitué de $(y_i, x_i)_{i=1,\cdots,n}$, n observations indépendantes du couple de vecteurs aléatoires (Y, X) de $\Re \times \Re^p$. On s'intéresse à la loi conditionnelle de $Y \mid X$.

Le même schéma directeur s'applique ici à partir des définitions des modèles et du critère (conditionnel) de Kullback-Leibler. Deux modèles sont proposés pour la modélisation de la densité conditionnelle de Y sachant X.

$$\mathcal{M}_1 \quad : \quad f(y_i \mid x_i, \beta) \quad ; \quad \beta \in \Theta_\beta$$

$$\mathcal{M}_2 \quad : \quad g(y_i \mid x_i, \gamma) \quad ; \quad \gamma \in \Theta_\gamma$$

Les log-vraisemblances conditionnelles associées à ces modèles sont[4]:

$$L_1(\beta) = \sum_{i=1}^n \log\ f(y_i \mid x_i, \beta) \quad et \quad L_2(\gamma) = \sum_{i=1}^n \log\ g(y_i \mid x_i, \gamma)$$

Nous pouvons introduire le critère (conditionnel) de Kullback-Leibler qui est ici :

[4]Les conditions usuelles de régularité sont supposées et ne seront pas détaillées dans cet exposé synthétique. Elles figurent par exemple dans l'ouvrage de Gourieroux et Monfort [37], volume 2, ou dans l'étude des pseudo-vraies valeurs réalisée par Dhaene [27].

$$E_\beta\left[\log\left(\frac{f(y_i \mid x_i,\beta)}{g(y_i \mid x_i,\gamma)}\right)\right] = \int_{\Re^d} \log\left(\frac{f(y_i \mid x_i,\beta)}{g(y_i \mid x_i,\gamma)}\right) f(y_i \mid x_i,\beta)dy_i$$

Ce critère diffère de celui donné dans la section précédente par le fait qu'il est conditionnel aux observations x_i.

Une mesure directionnelle de la distance entre $\mathcal{M}_1$ et $\mathcal{M}_2$ est :

$$\sum_{i=1}^{n} E_\beta\left[\log\ \frac{f(y_i \mid x_i,\beta)}{g(y_i \mid x_i,\gamma)}\right]$$

dont le minimum sur γ est réalisé par $\Gamma_n(\beta)$ qui est la "*pseudo-vraie valeur à distance finie*" de γ.

Il est à noter que $\Gamma_n(\beta)$ réalise, de manière équivalente, le maximum en γ de :

$$\sum_{i=1}^{n} E_\beta\left[\log\ g(y_i \mid x_i,\gamma)\right] \tag{1.5}$$

Lorsqu'on augmente la taille de l'échantillon,$(n \to \infty)$, la pseudo-vraie valeur à distance finie $\Gamma_n(\beta)$ tend vers $\Gamma(\beta)$ pseudo-vraie valeur (asymptotique) solution du problème de maximisation suivant :

$$\max_{\gamma\in\Theta_\gamma} \lim_{n\to\infty}\ \frac{1}{n}\sum_{i=1}^{n} E_\beta\left[\log\ g(y_i \mid x_i,\gamma)\right] = \max_{\gamma\in\Theta_\gamma} E_x E_\beta\left[\log\ g(y_i \mid x_i,\gamma)\right] \tag{1.6}$$

Où E_x désigne l'espérance relative à la distribution des x_i[5].

Remarque

La pseudo-vraie valeur à distance finie $\Gamma_n(\beta)$, maximum de l'expression (1.5), dépend donc des valeurs des variables (exogènes) x_i, et devrait être notée $\Gamma_n(\beta, X)$. Avant observation elle doit donc être considérée comme variable aléatoire. Par contre, la pseudo-vraie valeur asymptotique $\Gamma(\beta)$ issue de l'expression (1.6) n'est pas aléatoire, elle diffère donc par nature de $\Gamma_n(\beta)$. Ces deux notions sont toutefois confondues dans le cadre de modèles d'échantillonnage (où il n'y a pas d'exogènes) ainsi que dans les cas de modèles *iid*, ou autres modèles à valeur des x_i fixes ($f(y_i, x_i, \beta) = f(y_i, \beta)$). On parlera dans ces cas de pseudo-vraie valeur, sans distinction.

[5]On supposera, en effet, que la distribution empirique des x_i tend vers une distribution limite (et inconnue).

La pseudo-vraie valeur étant définie , nous pouvons introduire la notion d'enveloppement "*approché*", cette définition, plus familière, est essentiellement basée sur l'estimation de la différence d'enveloppement $\widehat{\gamma}-\Gamma(\widehat{\beta})$. L'expression de cette différence est centrale dans cette définition, elle servira de base aux tests d'enveloppement développés chapitre 2. La procédure de calcul de la pseudo-vraie valeur étant une minimisation, la transitivité de l'enveloppement exact ne se retrouvera pas dans l'enveloppement approché.

Afin d'être clair dans nos définitions nous parlerons d'*enveloppement* pour désigner l'enveloppement "*approché*" défini ici, l'enveloppement "*exact*" étant la dénomination réservée à la relation (1.1) de la définition 1.1.

1.2.4 Définition de l'enveloppement approché

Comme il n'est pas possible de vérifier la relation (1.1), l'idée est de définir l'enveloppement approché en se basant sur la différence entre l'estimateur $\widehat{\gamma}$ de γ dans $\mathcal{M}_2$ et un estimateur $\Gamma(\widehat{\beta})$ de la pseudo-vraie valeur $\Gamma(\beta)$, celle-ci ayant été calculée par minimisation de la spécificité.

Définition 1.3 *(Enveloppement approché) :*
On dira que "$\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ " ($\mathcal{M}_1\,\mathcal{E}\,\mathcal{M}_2$) si :

$$\widehat{\gamma}(Y_n) = \Gamma\left(\widehat{\beta}(Y_n)\right) \qquad (M_1 \quad p.s.) \tag{1.7}$$

$\Gamma\left(\widehat{\beta}(Y_n)\right)$ *étant l'estimateur de la pseudo-vraie valeur de γ sous $\mathcal{M}_1$.*

La différence fondamentale avec l'expression (1.1) définissant l'enveloppement exact, réside dans la connaissance de la pseudo-vraie valeur. Ici, elle est connue comme résultant d'une procédure de minimisation et l'on examine la nullité de la différence $\widehat{\gamma} - \Gamma(\widehat{\beta})$, contrairement à la définition de l'enveloppement exact où l'on s'intéressait à l'existence de Γ permettant la nullité de cette différence.

La relation (1.7) est évidemment dépendante de l'échantillon, et peut donc être testée. C'est d'ailleurs sur la différence $\widehat{\gamma} - \Gamma(\widehat{\beta})$, ou sur une fonction de cette différence que seront fondés les tests d'enveloppement classiques (voir chapitre 2).

Il est à noter que l'enveloppement approché n'est pas transitif, autrement dit, si $\mathcal{M}_1\,\mathcal{E}\,\mathcal{M}_2$ et $\mathcal{M}_2\,\mathcal{E}\,\mathcal{M}_3$ alors $\mathcal{M}_1$ n'enveloppe pas forcément le modèle $\mathcal{M}_3$. Cette situation est due au fait que les pseudo-vraies valeurs sont définies comme minimum d'un critère de "divergence" entre modèles qui n'est pas transitif (voir Dhaene [27]).

L'exemple suivant permet de représenter l'enveloppement approché sous une forme aussi simple que possible. Il est extrait de Hendry et Richard [54] et fait intervenir deux modèles univariés normaux non emboîtés.

Exemple 3 *$\mathcal{M}_1$ est un modèle proposant la densité de Y comme distribuée suivant une loi normale de variance unitaire, il est paramétré par la moyenne β et appartient donc à la famille de densités normales de variance 1.*

$$\mathcal{M}_1 \ : \ Y \sim \mathcal{N}(\beta, 1)$$

Ce modèle va s'opposer au modèle $\mathcal{M}_2$ proposant une distribution normale centrée paramétrée par sa variance γ^2.

$$\mathcal{M}_2 \ : \ Y \sim \mathcal{N}(0, \gamma^2)$$

Si $\beta \neq 0$ et $\gamma^2 \neq 1$ ces deux modèles sont non emboîtés, dans le sens où les familles paramétriques étudiées ici sont disjointes. Nous cherchons ici, à envelopper $\mathcal{M}_2$ par $\mathcal{M}_1$ en nous basant sur un échantillon $Y_n = (y_1, y_2, \ldots, y_n)$, de n réalisations indépendantes de la variable aléatoire réelle Y .

Les estimateurs associés aux paramètres de ces modèles sont :

- $\widehat{\beta} = \frac{1}{n} \sum_{i=1}^{n} y_i$, pour $\mathcal{M}_1$.
- $\widehat{\gamma^2} = \frac{1}{n} \sum_{i=1}^{n} y_i^2$, pour $\mathcal{M}_2$.

Ces estimateurs sont convergents dans leurs modèles respectifs. La pseudo-vraie valeur de γ^2 est elle obtenue par l'étude du comportement asymptotique de $\widehat{\gamma^2}$ sous $\mathcal{M}_1$. La décomposition suivante permet une analyse rapide du comportement asymptotique des différents termes.

$$\widehat{\gamma^2} = \frac{1}{n} \sum_{i=1}^{n} y_i^2 \ = \frac{1}{n} \left(\sum_{i=1}^{n} (y_i - \beta)^2 + 2\beta \cdot \sum_{i=1}^{n} (y_i - \beta) + \sum_{i=1}^{n} \beta^2 \right)$$

Sous $\mathcal{M}_1$

- le premier terme tend vers 1 (variance de y sous $\mathcal{M}_1$)
- le second s'annule (espérance d'une variable centrée)
- le dernier est exactement égal à β^2

Au total on obtient la pseudo-vraie valeur $\Gamma(\beta) = p\lim_{\mathcal{M}_1}\left(\widehat{\gamma^2}\right) = 1+\beta^2$

$\mathcal{M}_1$ enveloppera donc $\mathcal{M}_2$ *si et seulement si* $\widehat{\gamma}^2 = \Gamma(\widehat{\beta}) = 1+\widehat{\beta}^2$

Conformément à la définition de Hendry et Richard nous jugerons de l'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$, par la différence entre un estimateur de γ^2 et un estimateur de la pseudo-vraie valeur $\Gamma(\beta)$, donnant la statistique :

$$\widehat{\phi} = \widehat{\gamma^2} - \Gamma(\widehat{\beta}) = \widehat{\gamma}^2 - 1 - \widehat{\beta}^2 \tag{1.8}$$

basée sur l'échantillon Y_n.

En développant cette expression, la statistique s'écrit également sous la forme :

$$\widehat{\phi} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\beta})^2 - 1$$

Hendry et Richard nous proposent également d'examiner sur cet exemple la situation inverse où l'on cherche à tester l'enveloppement de $\mathcal{M}_1$ par $\mathcal{M}_2$.

La pseudo-vraie valeur associée à $\widehat{\beta}$ sous $\mathcal{M}_2$ est $\mathcal{B}(\gamma)$:

$$\mathcal{B}(\gamma) = p\lim_{\mathcal{M}_2}\widehat{\beta} = 0$$

$\mathcal{M}_2$ enveloppera donc $\mathcal{M}_1$ ssi $\widetilde{\phi} = \widehat{\beta}$ est nul . □

1.2.5 L'alternative de Gourieroux et Monfort

L'approche de Gourieroux et Monfort [39], [38] et [42], que nous qualifions d'alternative, présente la particularité de considérer explicitement le processus de génération des données comme extérieur aux modèles. Nous examinerons les possibilités d'enveloppement dans un sens ($\mathcal{M}_1$ enveloppe $\mathcal{M}_2$) comme dans l'autre, sans préférence a priori pour l'un des deux modèles. Ainsi, les deux modèles sont examinés de manière symétrique, le principe de l'enveloppement servant de critère de choix objectif.

Considérons le contexte conditionnel défini pour l'étude des pseudo-vraies valeurs, section 1.2.3.

Deux modèles sont proposés pour la modélisation de la densité conditionnelle de y sachant x.

$$\mathcal{M}_1 \quad : \quad f(y_i \mid x_i, \beta) \quad ; \quad \beta \in \Theta_\beta$$

$$\mathcal{M}_2 \quad : \quad g(y_i \mid x_i, \gamma) \quad ; \quad \gamma \in \Theta_\gamma$$

et supposons le processus de génération des données $\mathcal{P}_0$, extérieur aux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$. Il est caractérisé par la ("vraie") densité conditionnelle h :

$$\mathcal{P}_0 \quad : \quad h(y_i \mid x_i, \theta_0) \quad ; \quad \theta_0 \in \Theta_0$$

Puisque $\mathcal{P}_0$ est extérieur aux modèles, nous pouvons déterminer quel modèle $\mathcal{M}_1$ ou $\mathcal{M}_2$ est le plus proche de ce processus, au sens du contraste de Kullback et Leibler. Dans ce paysage, nous pouvons définir différentes pseudo-vraies valeurs selon le modèle de référence.

Si l'on prend pour modèle de référence le modèle $\mathcal{P}_0$, alors :

- la valeur minimisant la "distance" entre $\mathcal{P}_0$ et $\mathcal{M}_1$ est la pseudo-vraie valeur de β sous $\mathcal{P}_0$, β_0.
- celle minimisant la "distance" entre $\mathcal{P}_0$ et $\mathcal{M}_2$ est γ_0, la pseudo-vraie valeur de γ sous $\mathcal{P}_0$,

Ces valeurs sont définies comme solutions des programmes :

$$\begin{aligned} \beta_0 &= Arg\min_\beta \quad E_x E_0 \log\left[\tfrac{h(y_i|x_i,\theta_0)}{f(y_i|x_i,\beta)}\right] \\ &= Arg\max_\beta \quad E_x E_0 \left[\log\ f(y_i \mid x_i, \beta)\right] \\ et & \\ \gamma_0 &= Arg\max_\gamma \quad E_x E_0 \left[\log\ g(y_i \mid x_i, \gamma)\right] \end{aligned}$$

où E_x désigne l'espérance relative à la distribution des x_i et E_0 celle relative au "vrai" processus $\mathcal{P}_0$.

Malheureusement, en règle générale, le modèle $\mathcal{P}_0$ est inconnu et l'on ne peut donc pas choisir le modèle "*le plus proche*" au sens de ce critère.

On peut toutefois considérer l'un ou l'autre des modèles concurrents comme étant le "vrai" modèle. Si $\mathcal{M}_1$ est le modèle de référence, on définit la pseudo-vraie valeur $\Gamma(\beta)$ comme l'élément de Θ_γ minimisant la distance entre le modèle $\mathcal{M}_1$ au modèle $\mathcal{M}_2$. $\Gamma(\beta)$ est déterminée par le même type de maximisation :

$$\Gamma(\beta) = Arg\max_\gamma \ E_x E_\beta \log\ g(y_i \mid x_i, \gamma)$$

La fonction déterminant $\Gamma(\beta)$ dans Θ_γ, est la "*fonction de lien*" $\Gamma : \Theta_\beta \to \Theta_\gamma$, définie section 1.2.2.

Nous trouvons une expression symétrique en considérant $\mathcal{M}_2$ comme référence, la fonction de lien $\mathcal{B} : \Theta_\gamma \longrightarrow \Theta_\beta$, déterminant la pseudo-vraie valeur $\mathcal{B}(\gamma)$ dans Θ_β

$$\mathcal{B}(\gamma) = Arg\max_\beta \ E_x E_\gamma \log\ f(y_i \mid x_i, \beta)$$

Il est important de noter que les fonctions de lien Γ et $\mathcal{B}$ qui ne font intervenir que les modèles et leurs spécifications, sont indépendantes du vrai processus. N'ayant aucune hypothèse sur l'appartenance de ce processus à l'un des modèles, "*tout n'est donc que pseudo*".

Exemple 4 *Supposons les modèles conditionnels $\mathcal{M}_1$ et $\mathcal{M}_2$ linéaires gaussiens de variance unité :*

$$\mathcal{M}_1 \quad : \quad f(y \mid x, \beta) \quad = \quad \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\left(y - x_1'\gamma\right)^2}{2}\right)$$

et

$$\mathcal{M}_2 \quad : \quad g(y \mid x, \gamma) \quad = \quad \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\left(y - x_2'\gamma\right)^2}{2}\right)$$

Les fonctions de lien Γ et $\mathcal{B}$ vérifient :

$$\begin{aligned}
\Gamma(\beta) &= Arg\max_\gamma E_x E_\beta \left[\log\ g(y \mid x, \gamma)\right] \\
&= Arg\max_\gamma E_x E_\beta - (y - x_2'\gamma)^2 \\
&= Arg\max_\gamma E_x \left(x_1'\beta - x_2'\gamma\right)^2 \\
&= \left(E_x\left[x_2 x_2'\right]\right)^{-1} E_x\left[x_2 x_1'\right]\beta
\end{aligned}$$

De même :

$$\begin{aligned}
\mathcal{B}(\gamma) &= Arg\max_\gamma E_x E_\beta \left[\log\ g(y \mid x, \gamma)\right] \\
&= \left(E_x\left[x_1 x_1'\right]\right)^{-1} E_x\left[x_1 x_2'\right]\gamma
\end{aligned}$$

Les expressions définissant Γ et $\mathcal{B}$ sur cet exemple dépendent uniquement des spécifications des modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ (et notamment de la loi des x qui, toutefois, est souvent inconnue), et sont estimables en remplaçant les paramètres β et γ par leurs estimateurs $\widehat{\beta}$ et $\widehat{\gamma}$ dans les expressions ci-dessus. □

Dans ce contexte conditionnel, Gourieroux et Monfort [42] nous donnent leur définition de l'enveloppement, également proposée par Hendry et Richard [54], sous le terme "d'*enveloppement global*" ("*population encompassing*").

Définition 1.4 : *$\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ sous $\mathcal{P}_0$ ssi :*

$$\gamma_0 = \Gamma(\beta_0) \tag{1.9}$$

Cette définition de l'enveloppement fait ici intervenir explicitement le processus de génération des données $\mathcal{P}_0$, puisque les modèles sont représentés ici par l'intermédiaire de γ_0 et β_0, il est bien évident que la relation (1.9) ne lie pas

les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ dans l'absolu, c'est une relation liant les modèles pour "*un certain*" processus de génération des données $\mathcal{P}_0$ qui ne sera pas forcément vraie pour d'autres.

Propriété :

Il est intéressant de noter que si $\mathcal{P}_0 \in \mathcal{M}_1$, alors $\mathcal{M}_1$ enveloppe tout autre modèle $\mathcal{M}_2$.

Preuve :

Si $\mathcal{P}_0 \in \mathcal{M}_1$, alors $\forall\ \mathcal{M}_2$:

$$\begin{aligned} \gamma_0 &= Arg\max_\gamma\ E_x E_0\left[\log\ g(y_i \mid x_i, \gamma)\right] \\ &= Arg\max_\gamma\ E_x E_{\mathcal{M}_1}\left[\log\ g(y_i \mid x_i, \gamma)\right] \\ &= \Gamma(\beta_0) \end{aligned}$$

Donc $\mathcal{M}_1$ enveloppe le modèle $\mathcal{M}_2$. □

Les pseudo-vraies valeurs disponibles ici, et notamment $\Gamma(\beta)$ et $\mathcal{B}(\gamma)$, vont nous permettre de définir les "*ensembles images*" et "*ensembles réfléchis" :*

L'ensemble image de $\mathcal{M}_1$ dans $\mathcal{M}_2$ est[6] :

$$Im(\mathcal{M}_1) = \mathcal{M}_2' = \{g(y_i \mid x_i, \Gamma(\beta)),\ \beta \in \Theta_\beta\}$$

De même, l'image de $\mathcal{M}_2$ dans $\mathcal{M}_1$ est :

$$Im(\mathcal{M}_2) = \mathcal{M}_1' = \{f(y_i \mid x_i, \mathcal{B}(\gamma)),\ \gamma \in \Theta_\gamma\}$$

Les ensembles réfléchis $R_{\beta\gamma}$ et $R_{\gamma\beta}$ sont eux définis comme l'ensemble des points invariants par la double action des fonctions de liens, dans un sens et dans l'autre, soit plus formellement :

$$R_{\beta\gamma} = \{f(y_i \mid x_i, \beta)\ t.q.\ \beta = \mathcal{B}(\Gamma(\beta)),\ \beta \in \Theta_\beta\} \subset \mathcal{M}_1$$

et

$$R_{\gamma\beta} = \{g(y_i \mid x_i, \gamma)\ t.q.\ \gamma = \Gamma(\mathcal{B}(\gamma),\ \gamma \in \Theta_\gamma\} \subset \mathcal{M}_2$$

[6] M_1 peut être défini également comme $\{f(y_i \mid x_i, \beta)\ t.q.\ \beta \in \Theta_\beta\}$

Tout comme les fonctions de lien, ces ensembles sont définis dès que les modèles le sont, ils ne dépendent que de la forme des fonctions de lien et sont donc indépendants du processus de génération des données.

Exemple :

Si l'on se replace dans le cadre de l'exemple 4, on a :

$\mathcal{M}_2' = Im(\mathcal{M}_1)$ est le sous ensemble de $\mathcal{M}_2$ dont les paramètres γ appartiennent à l'image de la matrice $(E_x[x_2x_2'])^{-1}E_x[x_2x_1']$. □

Les ensembles présentés ci-dessus offrent, et c'est leur principal intérêt, d'importantes perspectives en vue de réduire les modèles en présence, en effet : Si $\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ alors le modèle image $\mathcal{M}_2'$ est à la même distance de $\mathcal{P}_0$ que $\mathcal{M}_2$. Autrement dit, $\mathcal{M}_1$ présente la même spécificité vis à vis de $\mathcal{P}_0$ que $\mathcal{M}_2$, de plus $\mathcal{M}_2' \subset \mathcal{M}_2$, il n'est donc pas nécessaire d'examiner le modèle $\mathcal{M}_2$ dans son intégralité. On peut ainsi réduire les modèles par examen des ensembles images.

Dans le cas d'enveloppement mutuel ($\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ et réciproquement), on peut remplacer les modèles initiaux par les ensembles images, puis par les images de ces ensembles images, etc... A la limite de ce processus on obtient les ensembles réfléchis $R_{\beta\gamma}$ et $R_{\gamma\beta}$ qui sont à la même distance de $\mathcal{P}_0$ que les modèles initiaux dont ils sont issus (voir Gourieroux et Monfort [39]), et qui au regard du principe de parcimonie, présentent un intérêt plus grand.

Un point important de cette étude repose sur les fonctions de lien, il faut noter que ces fonctions sont souvent inconnues, dans le cas de modèles avec variables exogènes, la distribution de ces variables est inconnue et les fonctions de lien, faisant intervenir cette distribution ne peuvent donc être déterminés explicitement. Un moyen de contourner cet obstacle est d'utiliser les fonctions de lien en échantillon fini définissant les pseudo-vraies valeurs finies de la section (1.2.3). Gourieroux et Monfort [40] proposent également une procédure de simulation de ces pseudo-vraies valeurs finies par tirages aléatoires d'éléments observés du processus (y_i, x_i).

1.3 Enveloppement parcimonieux et partiel

"The parcimony principle in empirical modelling is like Occam's razor : If a submodel has all the desirable properties of a larger model, we only need to consider the submodel."

Geert Dhaene (1993)

1.3.1 Enveloppement parcimonieux

Définition 1.5 *: $\mathcal{M}_1$ "enveloppe parcimonieusement" $\mathcal{M}_2$ si et seulement si :*

i) $\mathcal{M}_1$ enveloppe $\mathcal{M}_2$

ii) $\mathcal{M}_1$ est emboîté dans $\mathcal{M}_2$, au sens où $\mathcal{M}_1$ est un cas particulier de $\mathcal{M}_2$[7]

La totalité de l'information apportée par $\mathcal{M}_2$ se retrouve donc dans $\mathcal{M}_1$, ce qui peut constituer une importante avancée dans l'optique de réduire les modèles, et notamment dans une optique de prévision où la simplicité du modèle est souvent mise en avant.

Cette propriété présente de nombreux autres intérêts, en particulier comme le notent Hendry et Richard, les calculs des statistiques de test d'enveloppement sont simplifiés quand $\mathcal{M}_1$ est emboîté dans $\mathcal{M}_2$ (le calcul des pseudo-vraies valeurs y est en effet plus simple).

Nous verrons également chapitre 2 que les liens entre les tests d'enveloppement parcimonieux et les tests basés sur des conditions de moments (M-tests) s'avèrent être nombreux et étroits, comme le remarquent Lu et Mizon [61].

D'autre part, et plus fondamentalement, ce cadre permet l'étude du "*modèle emboîtant minimal*", c'est-à-dire du plus petit modèle $\mathcal{M}_c$ tel que $\mathcal{M}_1$ et $\mathcal{M}_2$ soient emboîtés dans $\mathcal{M}_c$. Intuitivement, il semble que ce modèle $\mathcal{M}_c$ ait la même spécificité que $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$. Lu et Mizon étudient les conditions pour lesquelles $\mathcal{M}_1$ enveloppe $\mathcal{M}_2$ *si et seulement si* $\mathcal{M}_1$ enveloppe $\mathcal{M}_c$, situation évidemment reliée à ce contexte d'enveloppement parcimonieux. L'exemple 3, présenté section 1.2.4, permet de visualiser aisément une telle situation.

Exemple 3 : (suite)

Les modèles en présence sont des modèles d'échantillonnages normaux, $\mathcal{M}_1$ appartient à la famille de densités normales de variance 1, $\mathcal{M}_2$ proposant une distribution normale centrée paramétrisée par sa variance.

$$\mathcal{M}_1 : Y \sim \mathcal{N}(\beta, 1) \quad et \quad \mathcal{M}_2 : Y \sim \mathcal{N}(0, \gamma^2)$$

[7] L'emboîtement peut être défini également, par inclusion des modèles, ou par inclusion des espaces paramétriques au sein d'une même famille de modèles, ou bien par la nullité du KLIC de $\mathcal{M}_2$ relativement à $\mathcal{M}_1$, ou par tout autre définition. Nous ne discuterons pas en détail de cette définition dans ce chapitre.

Le "*modèle minimal emboîtant*" $\mathcal{M}_1$ et $\mathcal{M}_2$ est défini ici par $\mathcal{M}_c$:

$$\mathcal{M}_c : Y \sim \mathcal{N}(m, v^2)$$

Les paramètres associés à ce modèle, $\delta = (m, v^2)$, sont estimés de manière convergente par $\widehat{\delta} = (\widehat{m}, \widehat{v}^2)$:

$$\widehat{m} = \widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}$$

$$\widehat{v^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

L'objectif étant de déterminer une condition pour que $\mathcal{M}_1$ enveloppe $\mathcal{M}_c$, nous nous intéressons à la pseudo-vraie valeur de δ sous $\mathcal{M}_1$, $\Delta(\beta) = (M(\beta), V^2(\beta))$ dont les expressions sont :

$$M(\beta) = \beta$$

$$V^2(\beta) = 1$$

En effet le comportement asymptotique de nos estimateurs sous $\mathcal{M}_1$ nous donne :

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} y_i \to \beta \text{ sous } \mathcal{M}_1$$

$$\widehat{v^2} = \sum_{i=1}^{n}(y_i - \overline{y})^2 \to 1 \text{ sous } \mathcal{M}_1$$

$\mathcal{M}_1$ enveloppe *parcimonieusement* $\mathcal{M}_c$ ssi $\widehat{\delta} = \Delta(\widehat{\beta})$ ce qui équivaut à:

$$\begin{cases} \widehat{\beta} = \widehat{\beta} \\ \widehat{v^2} = 1 \end{cases}$$

La première égalité est bien évidemment toujours vérifiée, la deuxième correspond à l'expression 1.8 définissant l'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$.

Un calcul rapide nous permet d'exprimer cette différence sous la forme :

$$\widehat{v^2} - 1 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\beta})^2 - 1 = \widehat{\gamma}^2 - 1 - \widehat{\beta}^2$$

Sur cet exemple, $\mathcal{M}_1$ enveloppe *parcimonieusement* $\mathcal{M}_c \Leftrightarrow \mathcal{M}_1$ enveloppe $\mathcal{M}_2$, les deux modèles $\mathcal{M}_c$ et $\mathcal{M}_2$ ont ainsi la même spécificité vis-à-vis de $\mathcal{M}_1$.

Nous reprendrons ultérieurement cet exemple dans l'optique de tester cette relation d'enveloppement en étudiant la distribution de $\widehat{v^2}$. □

1.3.2 Enveloppement partiel

Lu et Mizon[61] proposent en 93 des définitions plus générales en considérant la différence d'enveloppement, ou contraste, au travers d'une fonction pouvant être déterministe ou non. Nous donnons ici ces définitions d'enveloppement partiel, ou directionnel.

Définition 1.6 *(Enveloppement via une fonction) :*

$$\mathcal{M}_1 \textit{ enveloppe } \mathcal{M}_2 \textit{ via } C \quad \textit{ssi} \quad C(\gamma) - C(\Gamma(\beta)) = 0 \tag{1.10}$$

où C est une fonction connue, non aléatoire du paramètre γ de $\mathcal{M}_2$.

Ces auteurs proposent également une définition où l'on interprète la différence d'enveloppement par le biais d'une fonction tout-à-fait générale.

Définition 1.7 *(Lu et Mizon) :*

$$\mathcal{M}_1 \textit{ enveloppe } \mathcal{M}_2 \textit{ via } B \quad \textit{ssi} \quad E_{\mathcal{P}_0}\left[B(Y_n, \widehat{\gamma}) - E_{\mathcal{M}_1}\left[B(Y_n, \widehat{\gamma})\right]\right] = 0 \tag{1.11}$$

où :

- $E_{\mathcal{P}_0}$ désigne l'espérance relative au processus de génération des données $\mathcal{P}_0$
- B est une fonction des données Y_n et de $\widehat{\gamma}$.

Nous retrouvons dans cette expression les ingrédients de l'enveloppement approché :

- L'estimateur $\widehat{\gamma}$ est en fait généralisé par la fonction $B(Y_n, \widehat{\gamma})$ dans l'expression (1.11),
- tandis que $E_{\mathcal{M}_1}\left[B(Y_n, \widehat{\gamma})\right]$, qui est la réinterprétation de $B(Y_n, \widehat{\gamma})$ sous $\mathcal{M}_1$, remplace $\Gamma(\beta) = E_\beta\left[\widehat{\gamma}\right]$, la pseudo-vraie valeur de $\widehat{\gamma}$.

La différence est toutefois analysée ici sous $\mathcal{P}_0$.

Cette définition trouve sa source dans l'article de Mizon et Richard [66], l'introduction de la fonction B généralise la procédure d'estimation en quelque sorte, et permet d'élargir le champ d'action de l'enveloppement. Un grand nombre de statistiques de test peuvent ainsi être engendrées par différents choix de la fonction B.

Un exemple classique est la fonction B définie par :

$$B(Y_n, \widehat{\gamma}) = L_1(\widehat{\gamma}) - L_2(\widehat{\beta})$$

où $L_1(\widehat{\gamma})$ et $L_2(\widehat{\beta})$ désignent les log-vraisemblances des modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ respectivement.

L'hypothèse de test de l'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$ via B est alors la même que l'hypothèse du test de rapport de vraisemblance généralisé de Cox ([21] et [22]), que l'on trouve clairement explicité dans Pesaran [70]. Cette expression est également à la source de la notion d'enveloppement.

D'autres exemples de fonctions B peuvent être construits et sont étudiés par Mizon [65],Mizon et Richard [66] et Lu et Mizon [61] .

Il est à noter que cette dernière définition est la plus générale des définitions présentées ici et n'est pas reliée aux autres , alors que l'on constate que :

$$\mathcal{M}_1 \textit{ enveloppe exactement } \mathcal{M}_2 \Longrightarrow \mathcal{M}_1 \textit{ enveloppe } \mathcal{M}_2 \textit{ via } C$$

La réciproque de ce résultat n'est évidemment pas vraie puisqu'il est facile d'imaginer un modèle enveloppant partiellement un autre par construction d'une fonction C particulière, sans que la propriété d'enveloppement exact ne soit vérifiée.

Ces définitions sont des définitions d'enveloppement approché pour lesquelles la différence d'enveloppement est analysée au travers d'un filtre, la fonction B ou C, qui peut être directionnel, réducteur, ou généralisateur.

L'enveloppement partiel, proprement dit, est un cas particulier de ces définitions, correspondant à une définition de B (ou C) réduisant la dimension de $\widehat{\gamma}$ (une projection par exemple), seule "une partie" des paramètres de $\mathcal{M}_2$ est alors considérée comme pertinente pour l'analyse.

Nous verrons chapitre 2 que cette notion peut être génératrice de tests unifiant la littérature sur les tests de spécification, Lu et Mizon ajoutent même que tous les tests de spécification peuvent virtuellement être retrouvés par des choix appropriés des fonctions C et B .

1.4 Enveloppement bayésien

"In summary, encompassing is formalized as a concept of sufficiency among models whereas specificity mesures the lack of encompassing"

Jean-Pierre Florens, David F. Hendry et Jean-François Richard (1994)

1.4.1 Principe général

L'optique bayésienne propose d'associer aux modèles d'échantillonnage, des densités *a priori* sur les paramètres, permettant l'écriture d'une densité jointe à l'échantillon y et aux paramètres, dans chacun des modèles[8]. L'utilisation de la règle de Bayes pour décomposer cette densité jointe de deux façons différentes nous permet d'obtenir des densités *a posteriori* conditionnelles à l'échantillon y.

La loi jointe, $\pi(y, \beta)$, formée, sur $\mathcal{S} \times \Theta_\beta$, du produit de la densité d'échantillonnage par la densité *a priori* sur le paramètre, peut se décomposer également en la densité *a posteriori* que multiplie la densité prédictive (1.12).

Nous rappelons ici cette décomposition pour le modèle $\mathcal{M}_1$:

$$\begin{aligned} \pi(y, \beta) &= f(y \mid \beta) \cdot \mu(\beta) \\ &= \mu(\beta \mid y) \cdot f(y) \end{aligned} \tag{1.12}$$

Nous donnons dans la table ci-dessous les notations pour les deux modèles.

Notations bayésiennes

	Modèle 1	**Modèle 2**
Loi jointe	$\pi(y, \beta)$	$\pi(y, \gamma)$
Densité d'échantillonnage	$f(y \mid \beta)$	$g(y \mid \gamma)$
A priori	$\mu(\beta)$	$\nu(\gamma)$
A posteriori	$\mu(\beta \mid y)$	$\nu(\gamma \mid y)$
Prédictive	$f(y)$	$g(y)$

TeX fieldTOTeX field

L'objet de l'inférence bayésienne consiste à passer de l'*a priori* sur le paramètre à une densité *a posteriori* sur ce paramètre, conditionnellement aux observations. L'enveloppement bayésien se concentrera donc, tout naturellement, sur les densités *a posteriori*, les densités *a priori* n'étant de toutes façons pas comparables puisque basées sur des ensembles d'information différents.

1.4.2 Notion d'enveloppement bayésien

D'une manière similaire à la notion d'enveloppement classique, il y aura enveloppement bayésien de $\mathcal{M}_2$ par $\mathcal{M}_1$ si la densité *a posteriori* du modèle 1 explique celle du modèle 2 ou s'il existe une relation permettant de retrouver la densité *a posteriori* de $\mathcal{M}_2$ en utilisant celle de $\mathcal{M}_1$. Hendry et Richard

[8]Nous supposerons, dans toute cette partie, que les modèles bayésiens sont représentés par des densités, sans donner les conditions nécessaires à cette propriété.

[54] proposent une comparaison de l'*a posteriori* bayésien de $\mathcal{M}_2$ avec une interprétation de cet *a posteriori* par le modélisateur 1.

Définition 1.8 *(Enveloppement bayésien) :*
On dira que "le modèle bayésien $\mathcal{M}_1$ enveloppe $\mathcal{M}_2$", s'il existe une densité conditionnelle $\Gamma(\gamma \mid \beta)$ indépendante de y, telle que :

$$\nu(\gamma \mid y) = \int_{\Theta_\beta} \Gamma(\gamma \mid \beta)\mu(\beta \mid y)\ \partial\beta \tag{1.13}$$

"presque sûrement " en y.[9]

L'expression (1.13) exprime le fait que les résultats de $\mathcal{M}_2$, c'est-à-dire $\nu(\gamma \mid y)$, sont retrouvés à l'aide de ceux de $\mathcal{M}_1$, $\mu(\beta \mid y)$.

Le lien entre l'enveloppement classique et l'enveloppement bayésien réside dans la fonction Γ qui permettait de lier les espaces paramétriques et qui est remplacée ici par la densité de transition $\Gamma(\gamma \mid \beta)$ ou pseudo-vraie valeur bayésienne, liant les espaces de probabilité associés à chacun des modèles.

Une autre similitude avec l'enveloppement classique est que la relation (1.13) est rarement vérifiée, il faudra donc, ici encore, définir un critère de mesure du défaut d'enveloppement ou de mesure de la spécificité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$. Ce critère servira de base pour la détermination de la densité de transition $\Gamma(\gamma \mid \beta)$.

Auparavant, nous suivrons Florens, Hendry et Richard [31], sur la voie de la dualité existant entre l'enveloppement bayésien et l'exhaustivité entre statistiques au sens de Le Cam.

1.4.3 Enveloppement bayésien et exhaustivité

Rappelons tout d'abord la notion d'exhaustivité entre statistiques. Intuitivement, une statistique y est *exhaustive* pour une statistique z si y apporte la même information que z. Dans un contexte bayésien, l'information sur la loi de y sachant β est la même que celle sur z sachant β. Ou plus précisément :

Une statistique y est *exhaustive* pour une statistique z, conditionnellement au paramètre β, s'il existe une densité conditionnelle Λ indépendante de β, telle que :

$$g(z \mid \beta) = \int_{\mathcal{S}} f(y \mid \beta)\Lambda(z \mid y)\ \partial y \tag{1.14}$$

[9] Dans une optique bayésienne *"presque sûrement"* s'entend ici au sens de la loi prédictive de M_1, de densité $f(y)$.

où:

- $g(z \mid \beta)$ est la densité d'échantillonage de Z
- Λ est une densité conditionnelle sur z étant donné y, indépendante de β.

Selon Hendry et Richard, l'expression (1.13), qui relie deux paramètres (β, γ) et une statistique y, peut être exprimée de manière duale en introduisant deux statistiques (y, z) et un paramètre β. En effet, la substitution de (β, γ, y) par (y, z, β) mène immédiatement à l'expression (1.14)

L'enveloppement bayésien introduit ici, est ainsi réinterprété comme un concept *"d'exhaustivité entre modèles"* dual au concept *"d'exhaustivité entre statistiques"* défini par Le Cam [60].

Cette dualité ouvre des perspectives intéressantes en transposant les résultats connus dans le cadre de l'exhaustivité entre modèles, au cadre de l'enveloppement bayésien. La notion de *"déficience (deficiency) entre statistiques"*, comme mesure du manque d'exhaustivité, se retrouve notamment, dans la notion de la *"spécificité entre modèles"* comme mesure du défaut d'enveloppement bayésien.

La notion de spécificité résultant de cette étude duale permet l'introduction d'un critère pour la sélection de la densité de transition $\Gamma(\gamma \mid \beta)$.

1.4.4 Enveloppement bayésien et spécificité

Rappelons tout d'abord la définition d'une probabilité de transition dont est issue la densité $\Gamma(\gamma \mid \beta)$.

Définition 1.9 *Soient $(A, \mathcal{A})$ et $(C, \mathcal{C})$ deux espaces mesurables, une "probabilité de transition" est une fonction Λ :*

$$\Lambda : \quad \begin{array}{ccc} A \times \mathcal{C} & \longrightarrow & [0,1] \\ (a, Y) & \longrightarrow & \Lambda(a, Y) \end{array}$$

telle que :

i) $\forall\, a \in A,\ \Lambda(a, \cdot)$ *est une probabilité sur* $(C, \mathcal{C})$

ii) $\forall\, Y \in \mathcal{C},\ \Lambda(\cdot, Y)$ *est une fonction $\mathcal{A}$-mesurable.*

Comme dans le cas classique, raisonnons à densité de transition fixée $\Gamma(\gamma \mid \beta)$ afin de déterminer ensuite quel critère convient d'être utilisé pour la sélection de Γ.

Remarquons que la dualité construite ci-dessus, s'exprime par le passage d'un triplet (β, γ, y) à un autre triplet, "dual," (y, z, β). Nous pouvons construire sur le triplet (β, γ, y) une loi jointe π^* définie sur $\Theta_\beta \times \Theta_\gamma \times \mathcal{S}$, en utilisant cette densité de transition $\Gamma(\gamma \mid \beta)$[10].

$$\begin{aligned} \pi^*(\beta, \gamma, y) &= [f(y \mid \beta) \cdot \mu(\beta)]\, \Gamma(\gamma \mid \beta) \\ &= [f(y) \cdot \mu(\beta \mid y)]\, \Gamma(\gamma \mid \beta) \end{aligned}$$

La densité jointe π sur $\Theta_\beta \times \Theta_\gamma$ est ainsi une marginalisation de π^*.
Nous pouvons appliquer le même raisonnement sur π^* que sur π et appliquer la règle de Bayes de nouveau, afin de trouver l'*a posteriori* de γ (conditionnel à y) par :

$$\nu^*(\gamma \mid y) = \int_{\Theta_\beta} \mu(\beta \mid y) \Gamma(\gamma \mid \beta) d\beta \tag{1.15}$$

Nous trouvons ici l'interprétation personnelle par le propriétaire de $\mathcal{M}_1$ de la densité *a posteriori* sur γ, à partir de son propre *a posteriori* sur β. Dès lors nous sommes en présence de deux densités *a posteriori* sur γ sur la base desquelles peut s'effectuer la sélection de Γ.

Dans le même esprit que Le Cam , Hendry et Richard définissent la spécificité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$ par une mesure de la différence entre les deux densités *a posteriori* sur γ, $\nu^*(\gamma \mid y)$ et $\nu(\gamma \mid y)$.

Suivant les mesures choisies pour quantifier cette différence ou divergence (voir Hendry et Richard [54]), on obtiendra la spécificité, la p-spécificité, ou la φ-spécificité, comme minimum de la divergence espérée entre $\nu^*(\gamma \mid y)$ et $\nu(\gamma \mid y)$. Cette spécificité représente en fait la quantité incompressible séparant $\mathcal{M}_1$ de $\mathcal{M}_2$. Cette notion de divergence espérée minimale correspond à celle utilisée par Le Cam dans le contexte dual, pour mesurer le défaut d'exhaustivité entre statistiques.

[10] Florens et alii [31] suggèrent qu'il est naturel pour le propriétaire du modèle 1 de supposer que le paramètre β est suffisant pour caractériser la densité de y, c'est-à-dire de supposer l'indépendance de y et γ conditionnellement à β, soit en terme de densités :

$$f(y \mid \beta, \gamma) = f(y \mid \beta)$$

Pour compléter son information sur $\Theta_\beta \times \Theta_\gamma \times S$, le propriétaire de M_1 n'a besoin que d'une probabilité de transition de Θ_β sur Θ_γ

1.4.5 Enveloppement bayésien approché

Nous dressons ici un portrait semblable à celui rencontré dans le cadre de l'enveloppement classique. L'enveloppement défini par la relation (1.13) n'est que rarement vérifié, une procédure de mesure du défaut d'enveloppement est alors construite sur la spécificité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$. La pseudo-vraie valeur minimisant le contraste de Kullback-Leibler dans le cadre classique est ici remplacée par la "*transition optimale*" minimisant cette spécificité. La "*transition optimale*" Γ, ou pseudo-vraie valeur bayesienne est définie comme réalisant ce minimum sur une classe de densités de transition, malheureusement son calcul est souvent difficile, voire intraitable (voir Florens,Hendry et Richard [31]) Des méthodes de simulation sont toutefois capables de déterminer numériquement cette transition optimale, (voir Florens, Larribeau et Mouchart [33]), comme l'échantillonneur de Gibbs (voir Bouoiyour [13]). Une autre voie consiste à approcher la pseudo-vraie valeur, et à considérer l'enveloppement approché basé sur cette pseudo-vraie valeur.

Dans leur récent article sur l'enveloppement bayesien, Florens, Hendry et Richard, proposent trois solutions approchées du problème de minimisation déterminant la pseudo-vraie valeur :

- La première approximation consiste à reproduire la pseudo-vraie valeur classique, définie comme la $p\lim$ sous $\mathcal{M}_1$de l'estimateur du modèle $\mathcal{M}_2$, au cadre bayésien. Pour cela on considère la densité *a posteriori* de $\mathcal{M}_2$, $\nu(\gamma \mid y)$, dans l'optique de $\mathcal{M}_1$. La pseudo-vraie valeur approchée $\widetilde{\Gamma}(\gamma \mid \beta)$ est donc obtenue comme marginalisation de l'*a posteriori* $\nu(\gamma \mid y)$ de $\mathcal{M}_2$, par rapport à l'échantillon, en utilisant la densité d'échantillonnage de $\mathcal{M}_1$, $f(y \mid \beta)$:

$$\widetilde{\Gamma}(\gamma \mid \beta) = \int \nu(\gamma \mid y) \cdot f(y \mid \beta) dy$$

  En décomposant la densité d'échantillonnage $f(y \mid \beta)$, on obtient :

$$\widetilde{\Gamma}(\gamma \mid \beta) = \frac{1}{\mu(\beta)} \int \nu(\gamma \mid y) \cdot \mu(\beta \mid y) f(y) dy$$

  Cette densité de transition, bien que n'étant pas la transition optimale, présente l'avantage d'être facilement calculable dans un grand nombre d'applications[11].

- La deuxième approximation consiste à définir la densité de transition comme la fonction linéaire minimisant un critère de moindres carrés :

[11] En effet, dans le cas où les deux modèles présentent les mêmes densités a priori et d'échantillonnage, la densité de transition $\widetilde{\Gamma}(\gamma \mid \beta)$ se retrouve réduite à une Dirac.

Soient β et γ des variables aléatoires de dimensions l et m, notons $\widehat{\beta}$ et $\widehat{\gamma}$ les espérances $E_1[\beta \mid y]$ et $E_2[\gamma \mid y]$, respectivement. La pseudo-vraie valeur de γ relative à β est la fonction linéaire $\widehat{\Gamma}(\gamma \mid \beta) = \widehat{\Lambda}'\beta$ où $\widehat{\Lambda}$ est la matrice minimisant :

$$E_1\left[\left(\Lambda'\widehat{\beta} - \widehat{\gamma}\right)'\left(\Lambda'\widehat{\beta} - \widehat{\gamma}\right)\right]$$

qui est solution du système :

$$E_1\left(\widehat{\beta}\widehat{\beta}'\right)\widehat{\Lambda} = E_1\left(\widehat{\beta}\widehat{\gamma}'\right)$$

Si $E_1\left(\widehat{\beta}\widehat{\beta}'\right)$ est non singulière alors $\widehat{\Lambda}$, (et donc $\widehat{\Gamma}(\gamma \mid \beta)$), est unique.

Cette approximation possède ainsi les charmes d'un calcul aisé et d'un comportement asymptotique agréable puisque, sous des conditions techniques non reproduites ici, on a :

Si $\widehat{\gamma}$ converge vers $\gamma(\beta)$ sous la loi jointe de $\mathcal{M}_1$ alors :

$$\widehat{\Lambda} \to \left[E_1(\beta\beta')\right]^{-1} E_1(\beta\gamma'(\beta))$$

- Une dernière technique consiste à partitionner les espaces paramétriques Θ_β et Θ_γ en un nombre fini de sous espaces $(\Theta_i)_{i=1,.m}$ et $(\Theta_j)_{j=1,.,m}$ respectivement.

  Une pseudo-vraie valeur discrète est alors proposée comme la matrice $\Delta = (\delta_{i,j})$ dont chaque élément est défini comme permettant la minimisation de la spécificité totale décomposée sur les sous espaces. Par exemple, si le critère de spécificité utilisé est le KLIC, on obtient la pseudo-vraie valeur discrète Δ comme solution de :

$$\min_{\delta_{i,j}} E_1\left\{\sum_{i,j}\delta_{i,j}\mu(\Theta_i \mid \mathcal{S})\log\left[\frac{\sum_i \mu(\Theta_i \mid \mathcal{S})}{\nu(\Theta_i \mid \mathcal{S})}\right]\right\}$$

  L'espérance E_1 est ici relative à $f(y)$ des techniques d'évaluations sont également proposées par les auteurs qui précisent que cette transition discrète peut être rendue arbitrairement proche de la transition optimale en augmentant la taille de la partition. L'enveloppement bayésien approché est alors basé sur la nullité de la spécificité estimée, une fois la pseudo-vraie valeur déterminée, les tests sont également basés sur ces expressions.

1.5 Conclusion

"An essential characteristic of empirical modelling (and in fact in the developpement of theory models) is that it is not a "once-for all" event, but a process in which new information from theory and/or data leads to modification of existing models. It seems reasonable to require, therefore, that this process be progressive rather than degenerate and use of encompassing principle helps to ensure this."

Grayham E. Mizon (1984)

La notion d'enveloppement, que nous venons de détailler, se fonde sur l'existence d'une fonction de lien permettant l'interprétation des résultats d'un modèle $\mathcal{M}_1$ par ceux d'un autre modèle $\mathcal{M}_2$. Cette relation exacte, formelle, est transitive, et relie les modèles par l'intermédiaire des estimateurs qui leur sont associés. C'est par l'existence de cette fonction que les idées de progressivité dans la validation de nouveaux modèles et de comparaison stratégique de modèles, ont été formalisées.

L'existence, ou la non-existence, d'une fonction étant difficile à assurer, ce principe trouve naturellement son application dans la notion approchée de l'enveloppement. La pseudo-vraie valeur nous donne en effet une possibilité de lier les espaces paramétriques associés aux modèles. Il ne s'agit plus alors de "*trouver*" la fonction de lien mais de "*vérifier*" que la pseudo-vraie valeur est "*suffisamment proche*" de l'estimateur associé à $\mathcal{M}_2$. Contrairement à l'approche symétrique de Gourieroux et Montfort prenant explicitement en compte le "*vrai*" modèle, notre approche est directionnelle puisque nous construisons la différence d'enveloppement, ainsi que la pseudo-vraie valeur, sur la base d'un modèle de référence $\mathcal{M}_1$. C'est sur la différence entre estimateur et pseudo-vraie valeur que vont être fondés les tests d'enveloppement, que nous développons dans le chapitre suivant.

[a4,12pt]freport

Chapter 2

Tests d'enveloppement

"Most empirical testing is to ascertain the status of empirical models, not to test theories. However, here again encompassing helps resolve the problem."
David F. Hendry (1993)

2.1 Que teste-t-on ?

Les tests présentés ici sont directement issus des grands principes d'inférence classiques : le principe du rapport de vraisemblance, le principe de Wald, et le principe du Score. Les tests sont asymptotiques par nature, et ont été introduits principalement par Mizon et Richard [66], Gourieroux et Monfort [39], et Florens, Hendry et Richard [31] dans un contexte bayésien. Il importe toutefois d'être précis sur l'hypothèse que l'on cherche à tester, ainsi que sur le modèle pris pour référence lors de ces tests. Dans leur étude, Hendry et Richard [54], distinguent d'ailleurs deux approches de l'enveloppement, selon le modèle de référence. Ces auteurs distinguent ainsi le "*sampling encompassing*" du "*population encompassing*", selon que la différence d'enveloppement est examinée sous l'optique du modèle $\mathcal{M}_1$, ou sous celle du processus de génération des données $\mathcal{P}_0$.

Gourieroux et Montfort [38], se placent sous la direction du "vrai" processus de génération des données $\mathcal{P}_0$ et étudient l'hypothèse nulle :

$$H_0 \quad : \quad \gamma_0 = \Gamma(\beta_0)$$

Sous H_0, la limite de $\widehat{\phi_0} = (\widehat{\gamma_0} - \Gamma(\widehat{\beta_0}))$ entre les estimateurs des pseudo-vraies valeurs γ_0 et $\Gamma(\beta_0)$ tend vers zéro, un test de Wald est alors défini

(WET), ainsi qu'un Test du Score (SET), enfin un test d'enveloppement généralisé (GET), est également proposé.

Mizon et Richard [66], étudient l'enveloppement sous l'optique du modèle $\mathcal{M}_1$, l'hypothèse nulle, directement issue de l'enveloppement exact est alors :

$$H_1 \quad : \quad \gamma = \Gamma(\beta)$$

L'enveloppement exact ne pouvant, par nature, être testé, c'est l'enveloppement approché qui sert donc de base à ces tests, on va donc tester la nullité de la spécificité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$, la pseudo-vraie valeur ayant été préalablement déterminée. La statistique de test sera alors basée sur la différence entre un estimateur du paramètre du second modèle et un estimateur de la pseudo-vraie valeur. Pour cela on retrouve les deux grandes orientations classiques :

- Tester directement la nullité de $\widehat{\phi}$ défini comme la différence $\widehat{\gamma} - \Gamma(\widehat{\beta})$, ou d'une fonction de $\widehat{\phi}$ (tests de Wald)
- Tester la nullité du score, i.e. dériver le critère (test du Score)

En règle générale la distribution de $\widehat{\phi}$ n'est pas connue en échantillon fini, et il est alors nécessaire d'avoir recours à une étude asymptotique pour caractériser la distribution de $\widehat{\phi}$.

L'exemple 3 permet une approche simple des tests d'enveloppement, sur cet exemple la distribution de $\widehat{\phi}$ sera aisément caractérisée.

Exemple 3 (suite et fin) :

Le modèle $\mathcal{M}_1 : Y \sim \mathcal{N}(\beta, 1)$ enveloppe le modèle $\mathcal{M}_2 : Y \sim \mathcal{N}(0, \gamma^2)$, si la condition :

$$\widehat{v^2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{\beta})^2 = 1$$

est vérifiée.

L'originalité du modèle $\mathcal{M}_2$ par rapport à $\mathcal{M}_1$ consiste, en effet, à laisser la variance libre alors qu'elle est contrainte à 1 dans $\mathcal{M}_1$. Le test d'enveloppement portera donc, logiquement, sur l'égalité à 1 de la variance de Y, sous $\mathcal{M}_1$, c'est-à-dire sur la pertinence de cette originalité de $\mathcal{M}_2$ vis-à-vis de $\mathcal{M}_1$.

Sous $\mathcal{M}_1$ la distribution de $\widehat{v^2}$ est connue puisqu'il s'agit d'une distribution de Khi-deux.

$$\widehat{v^2} \underset{\sim}{\overset{\mathcal{M}_1}{}} \frac{1}{n-1}\,\chi^2_{(n-1)}$$

La situation inverse où l'on cherche à tester l'enveloppement de $\mathcal{M}_1$ par $\mathcal{M}_2$ nous donne également une distribution de Khi-deux. Les rôles étant inversés, le modèle pris pour référence est maintenant $\mathcal{M}_2$, c'est donc sous $\mathcal{M}_2$ que l'on examine la pseudo-vraie valeur et la distribution de la statistique de test.
La pseudo-vraie valeur associée à $\widehat{\beta}$ sous $\mathcal{M}_2$ est $\mathcal{B}(\gamma) = 0$:

$\mathcal{M}_2$ enveloppera donc $\mathcal{M}_1$ ssi $\widetilde{\phi} = \widehat{\beta} - \mathcal{B}(\widehat{\gamma}) = \widehat{\beta}$ est nul sous $\mathcal{M}_2$.

La distribution de $\widetilde{\phi} = \widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ sous $\mathcal{M}_2$, est évidemment une distribution normale centrée.

$$\widetilde{\phi} \overset{\mathcal{M}_2}{\sim} \mathcal{N}(0, \frac{\gamma^2}{n})$$

Nous obtenons ainsi la statistique de test de Wald suivante :

$$\eta = \frac{n \cdot \widehat{\phi}^2}{\widehat{\gamma}^2} \overset{\mathcal{M}_2}{\rightsquigarrow} \chi^2_{(1)}$$

□

2.1.1 Tests de Wald (Wald Encompassing Tests)

Afin de construire un test de l'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$, Mizon et Richard [66], suivant les travaux de Cox [21] et [22], Huber [55] et White [91], nous donnent la distribution limite sous $\mathcal{M}_1$ de $\sqrt{n} \cdot \widehat{\phi} = \sqrt{n}(\widehat{\gamma} - \Gamma(\widehat{\beta}))$.

Ici la pseudo-vraie valeur $\Gamma(\beta)$ est définie comme $E_\beta(\widehat{\gamma})$, où E_β désigne l'espérance sous $\mathcal{M}_1$, si nécessaire, on remplacera cette espérance par la $p\lim$ sous $\mathcal{M}_1$ de l'estimateur $\widehat{\gamma}$. Nous donnons ces résultats dans une version allégée, sans démonstration ni hypothèses précises, laissant le soin au lecteur de se reporter aux textes originaux pour plus de précision.

Théorème 2.1 *Sous les "conditions usuelles de régularité du maximum de vraisemblance" (voir White [91], conditions A1-A7), la distribution jointe des estimateurs du maximum de vraisemblance $\widehat{\beta}$ et $\widehat{\gamma}$, sous $\mathcal{M}_1$ est :*

$$\sqrt{n}\begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} - \begin{pmatrix} \beta \\ \Gamma(\beta) \end{pmatrix} \overset{\mathcal{M}_1}{\rightsquigarrow} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_\beta(\widehat{\beta}) & V_\beta(\widehat{\beta}) \cdot D' \\ D \cdot V_\beta(\widehat{\beta}) & V_\beta(\widehat{\gamma}) \end{pmatrix}\right)$$

où $V_\beta(\widehat{\beta})$ est la matrice de variance-covariance usuelle pour l'estimateur du maximum de vraisemblance d'un modèle correctement spécifié, tandis que $V_\beta(\widehat{\gamma})$ est celle de l'estimateur du maximum de vraisemblance d'un modèle mal-spécifié.

Soit, si $L_i(\beta)$ désigne la vraisemblance associée au modèle $\mathcal{M}_i$:

- $V_\beta(\widehat{\beta}) = \left(\lim_{n\to\infty} -\frac{1}{n}\, E\, \frac{\partial^2 L_1(\beta)}{\partial\beta\partial\beta\prime}\right)^{-1}$
- $V_\beta(\widehat{\gamma}) = HJH$

  avec :

  $J = \lim_{n\to\infty}\ E_\beta\ \left(\frac{1}{n}\frac{\partial L_2(\gamma)}{\partial\gamma}\ \ \frac{\partial L_2(\gamma)}{\partial\gamma\prime}\right)$

  $H = \left(\lim_{n\to\infty}\ -\frac{1}{n}E_\beta\ \left[\frac{\partial^2 L_2(\gamma)}{\partial\gamma\partial\gamma\prime}\right]\right)^{-1}$

  et D est la matrice constituée des dérivées $\left(\frac{\partial\Gamma(\beta)}{\partial\beta\prime}\right)$.

La distribution limite de $\widehat{\phi}$ découle de cette expression et l'on a :

$$\sqrt{n}\cdot\ \widehat{\phi}\ \overset{\mathcal{M}_1}{\rightsquigarrow}\ \mathcal{N}\left(0, V_\beta(\widehat{\phi})\right) \tag{2.1}$$

avec $V_\beta(\widehat{\phi}) = V_\beta(\widehat{\gamma}) - DV_\beta(\widehat{\beta})D\prime$

Une statistique de test de Wald est maintenant construite sur la base de la distribution limite de $\sqrt{n}\cdot\widehat{\phi}$ sous $\mathcal{M}_1$.

Corollaire 2.2 : *Sous les hypothèses du théorème précédent, un test de Wald de l'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$ est donné par la statistique :*

$$\eta_1 = n\cdot\widehat{\phi}\prime V_\beta(\widehat{\phi})^+\widehat{\phi}$$

$V_\beta(\widehat{\phi})$ n'étant pas toujours inversible, $V_\beta(\widehat{\phi})^+$ désigne un inverse généralisé de $V_\beta(\widehat{\phi})$, on note l son rang.

La statistique η_1 a alors une distribution limite de $\chi^2_{(l)}$ sous $\mathcal{M}_1$.

2.1.2 Test du Score (Score Encompassing Test)

Le test du score est basé sur la dérivée de la vraisemblance du modèle $\mathcal{M}_2$, estimé pour la pseudo-vraie valeur $\Gamma(\widehat{\beta})$, pour cela définissons le score par S :

$$S(\gamma) = \frac{1}{\sqrt{n}}\cdot\frac{\partial L_2(\gamma)}{\partial\gamma}$$

La statistique du score est basée sur la nullité de $S\left(\Gamma(\widehat{\beta})\right)$. Par définition, $S(\widehat{\gamma}) = 0$, en développant S au voisinage de $\widehat{\gamma}$ on a :

$$S(\widehat{\gamma}) = 0 = S(\Gamma(\widehat{\beta})) - \sqrt{n}H\cdot(\Gamma(\widehat{\beta}) - \widehat{\gamma}) + o_p(1)$$

soit encore :

$$S(\Gamma(\widehat{\beta})) = \sqrt{n}H \cdot \widehat{\phi} + o_p(1) \tag{2.2}$$

où la matrice H (supposée régulière) est définie par :

$$H = \lim_{n\to\infty} \left[-\frac{1}{n} E_\beta \frac{\partial^2 L_2(\gamma)}{\partial\gamma\partial\gamma'} \right]_{\gamma=\Gamma(\widehat{\beta})}$$

L'équation (2.2) suggère l'utilisation de la formule (2.1) pour définir la statistique de test du score, η_2, par :

$$\eta_2 = S(\Gamma(\widehat{\beta}))' V_\beta(S)^+ S(\Gamma(\widehat{\beta}))$$

où $V_\beta(S)^+$ désigne une inverse généralisée de $V_\beta(S)$:

$$V_\beta(S) = H\ V_\beta(\widehat{\phi})\ H$$

Nous renvoyons à Mizon et Richard [66], pour plus de détails concernant les hypothèses sous lesquelles ces tests sont établis, ainsi que celles assurant de l'équivalence asymptotique du test du Score et du test de Wald. Ces auteurs sont également à l'origine du développement de tests plus généraux basés sur l'utilisation de l'enveloppement étudié via une fonction B.

2.1.3 Tests classiques et enveloppement.

Lu et Mizon [61], mettent également en évidence les relations entre les tests d'enveloppement et les tests classiques par l'utilisation judicieuse de l'expression (1.11) définissant l'enveloppement via une fonction B (voir section 1.3.1). Afin de généraliser et d'étendre la notion d'enveloppement, Mizon et Richard [66] proposaient, en effet, de s'intéresser à $\widetilde{\gamma} = B(Y_n, \widehat{\gamma})$ et nous donnent, dans le théorème suivant, la distribution asymptotique de la statistique $\widetilde{\phi} = \widetilde{\gamma} - E_{\mathcal{M}_1}[\widetilde{\gamma}]$.

Théorème 2.3 *Sous les hypothèses du théorème 2.1, et sous les hypothèses de régularité de B et de $K = p\lim_{\mathcal{M}_1}\left(\frac{\partial B(Y_n,\gamma)}{\partial\gamma\prime}\right)$ énoncées par Mizon et Richard [66](voir annexe), on a :*

$$\sqrt{n}\widetilde{\phi} = \sqrt{n}\ (\widetilde{\gamma} - E_{\mathcal{M}_1}[\widetilde{\gamma}]) \overset{\mathcal{M}_1}{\rightsquigarrow} \mathcal{N}\left(0, V_\beta(\widetilde{\phi})\right) \tag{2.3}$$

où : $V_\beta(\widetilde{\phi}) = KV_\beta(\widehat{\phi})K\prime$

Corollaire 2.4 : *Le test de Wald associé à* $\widetilde{\phi}$ *est donné par :*

$$\eta_1' = n\widetilde{\phi}' V_\beta(\widetilde{\phi})^+ \widetilde{\phi} \tag{2.4}$$

où : r et $V_\beta(\widetilde{\phi})^+$ *désignent le rang et une inverse généralisée de* $V_\beta(\widetilde{\phi})$.

La statistique η_1' *a alors une distribution limite de* $\chi^2_{(r)}$ *sous* $\mathcal{M}_1$.

Le test du Score peut également être retrouvé par la même procédure que section (2.1), et Mizon et Richard [66], montrent que ce test est asymptotiquement équivalent au test de Wald sous la condition que $\sqrt{n}\phi^*$ soit asymptotiquement négligeable ($o_p(1)$), avec :

$$\begin{aligned} \phi^* &= B(Y_n, \Gamma(\beta)) - E_{\mathcal{M}_1}\left[B\left(Y_n, \Gamma(\beta)\right)\right] - \phi_\beta\left(\beta - \widehat{\beta}\right) \\ et & \\ \phi_\beta &= \lim_{n\to\infty} \tfrac{1}{n} E_\beta\left(B(Y_n, \Gamma(\beta)) \cdot \tfrac{\partial L_1}{\partial \beta}\right) \end{aligned}$$

La classe de tests définis par (2.3) et (2.4) permet de retrouver un bon nombre de tests classiques en économétrie , l'exemple le plus célèbre est celui du test de Cox obtenu par un choix judicieux de la fonction B.

En effet, si l'on choisit $B(Y_n, \widehat{\gamma}) = \frac{1}{n}\left(L_1(\widehat{\gamma}) - L_2(\widehat{\beta})\right)$ où $L_1(\widehat{\gamma})$ et $L_2(\widehat{\beta})$ désignent les log-vraisemblances des modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ respectivement, on obtient η_1' comme étant la statistique du rapport de vraisemblance généralisé de Cox.

Un autre exemple est le critère d'information de Sawyer (83) cité par Mizon [65] qui est retrouvé en posant $B(Y_n, \widehat{\gamma}) = \frac{1}{n} E_{\widehat{\gamma}}\left(L_1(\widehat{\gamma}) - L_2(\widehat{\beta})\right)$ qui est un estimateur du critère d'information de Kulback et Leibler entre $\mathcal{M}_1$ et $\mathcal{M}_2$. Ce test présente l'avantage de n'être pas soumis à la condition d'orthogonalité du test de Cox. Mizon [65] propose de nombreux exemples de tests pouvant être retrouvés ainsi[1].

2.2 Enveloppement et choix de régresseurs paramétriques

Le problème du choix des régresseurs constitue l'un des problèmes majeurs de l'économétrie depuis de longues années. De nombreuses procédures de sélection ont été proposées dans le cadre classique (voir Pesaran [70]), dans le cadre bayésien, (voir Zellner [93]) que les modèles soient spécifiés paramétriquement ou non-paramétriquement (voir la synthèse de Lavergne [58]).

[1] Parmi ceux-ci, on trouve le test pour déceler la présence de facteurs communs dans les processus autorégressifs (COMFAC) de Sargan (64), ainsi que le test directionnel de Epps et ali.(82)

2.2.1 Modèle de régression

Soit (X, Y) un vecteur aléatoire défini sur l'espace mesuré $(\Re^p \times \Re, \mathcal{B}_{\Re^{p+1}}, \lambda)$, nous supposerons que ce couple admet la densité[2] $\varphi(x, y)$ par rapport à la mesure de Lebesgue λ.

La régression de Y sur X s'écrit alors mathématiquement :

$$f(x) = E[Y \mid X = x] = \frac{\int y\varphi(x, y)\lambda(dy)}{\int \varphi(x, y)\lambda(dy)} = \frac{\int y\varphi(x, y)\lambda(dy)}{\varphi(x)}$$

en tout point où $\varphi(x)$ est non nulle.

On trouve souvent le modèle de régression sous la forme :

$$Y = f(X) + U \tag{2.5}$$

où $E[U \mid X] = 0$, λ-presque sûrement.

Il est important de lire cette expression dans le bon sens. Ici l'équation (2.5) se lit de "*la gauche vers la droite*" puisque la partie gauche détermine le résidu U intervenant dans la partie droite, U est donc défini par :

$$U = Y - f(x) \tag{2.6}$$

Remarque :

Hendry [53] nous rappelle que l'on trouve souvent formulés identiquement deux concepts totalement différents sous une équation du type :

$$y_i = f(x_i) + \eta_i$$

Si l'on a affaire à une "*expérience contrôlée*", y_i est le résultat de la $i$ème expérience, x_i est la variable d'entrée, f est la fonction liant les deux et η_i est une perturbation qui varie entre les expériences. Cette équation se lit de "*la droite vers la gauche*" puisque pour le même input x_i, on retrouvera (modulo la perturbation) le même output y_i. C'est ainsi notamment que doit se concevoir l'idée d'un "*vrai*" modèle, tel que le processus de génération des données $\mathcal{P}_0$.

En économétrie par contre, les modèles sont des approximations de la réalité ; y_i est engendré par un processus inconnu que l'on cherche à "mimer", on le décompose alors en une partie explicative $f(x_i)$ et une partie inexpliquée η_i définie comme :

$$\eta_i = y_i - f(x_i)$$

[2]Comme précédemment nous noterons les densités marginales et conditionnelles par la même fonction φ, les arguments de cette fonction levant toute ambiguïté.

Des changements dans la modélisation entraînent donc des changements pour η, l'équation se lit ainsi de *"la gauche vers la droite"*.

La régression linéaire est une approximation de la réalité pour laquelle on impose une spécification particulière de f et de η, ce modèle est ainsi un modèle approché du modèle de régression exact défini par (2.5). □

La régression linéaire est donc présentée comme une spécification de la fonction f contrainte à être linéaire en X, la distribution des résidus peut aussi être spécifiée pour donner le "*modèle linéaire normal* ". Enfin, si les résidus sont de plus supposés indépendants et de même variance, on obtient le "*modèle linéaire standard*".

Il arrive souvent que l'on veuille sélectionner des modèles de régression en choisissant entre des ensembles de régresseurs définissant des modèles non emboîtés. Nous entendons par "non emboîtés" des modèles tels qu'aucun des deux modèles ne peut s'exprimer comme une particularisation ou une généralisation de l'autre.

Cette question de la sélection de régresseurs a donné lieu à de nombreux travaux en économétrie, voir entre autre Amemiya [2], Atkinson [4], Hausman [52] ou Pesaran [70]. Le problème du choix de régresseurs dans le cadre de la régression linéaire normale a notamment été longuement étudié. Nous pouvons introduire ce problème de choix de modèles tel qu'il se présente généralement en économétrie.

Soit $S_i = (Y_i, X_i, Z_i)_{i=1,\dots,n}$, n réalisations indépendantes du vecteur aléatoire S de $\Re \times \Re^p \times \Re^q$. Essentiellement, X et Z représentent les variables exogènes associées aux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$.

Le problème s'écrit généralement sous la forme :

$$\begin{aligned} \mathcal{M}_1 : \quad & y = X\beta + u \qquad u \sim \mathcal{N}(0, \sigma^2 I_n) \\ \mathcal{M}_2 : \quad & y = Z\gamma + v \qquad v \sim \mathcal{N}(0, \tau^2 I_n) \end{aligned} \tag{2.7}$$

où X et Z représentent les matrices de régresseurs de dimensions $(n \times p)$ et $(n \times q)$ respectivement, et où y est un vecteur d'observations de dimension $(n \times 1)$.

En fait, ce problème peut se présenter de différentes manières et Mizon [65] nous met en garde sur la modélisation qui en est faite. On peut, en effet, réinterpréter le système (2.7) comme la donnée de deux modèles conditionnels, l'un par rapport à la variable X, l'autre par rapport à Z.

$$\begin{aligned} \mathcal{M}_1 : \quad & y \mid X \sim \mathcal{N}(X\beta, \sigma^2 I_n) \\ \mathcal{M}_2 : \quad & y \mid Z \sim \mathcal{N}(Z\gamma, \tau^2 I_n) \end{aligned} \tag{2.8}$$

Toutefois cette interprétation présente l'inconvénient de séparer complètement les modèles $\mathcal{M}_1$ et $\mathcal{M}_2$, ces deux modèles reposant sur des distributions conditionnelles complètement différentes. Cette formulation n'est donc pas satisfaisante, d'autant que le modèle $\mathcal{M}_1$ "*ne dit rien*" sur la variable Z, les deux modèles dans (2.8) pouvant également être simultanément acceptés si y,X et Z ont une distribution jointe normale multivariée, par exemple.

Une approche permettant d'introduire une distribution commune et donc des hypothèses susceptibles d'être testées est :

$$\begin{aligned} \mathcal{M}_1 : \quad y \mid X, Z \quad &\sim \mathcal{N}(X\beta, \sigma^2 I_n) \\ \mathcal{M}_2 : \quad y \mid Z, X \quad &\sim \mathcal{N}(Z\gamma, \tau^2 I_n) \end{aligned} \tag{2.9}$$

La formulation (2.9) indique que nous avons deux modèles conditionnels aux mêmes variables (X, Z), et donc relatives à la même distribution, et précise que le modèle $\mathcal{M}_1$ exclut la variable Z de la modélisation, tandis que $\mathcal{M}_2$ exclut la variable X, ce qui nous donne généralement des modèles non-emboîtés. Nous nous efforcerons de garder cette formulation du problème tout au long de ce travail.

Dans l'étude non-paramétrique à venir le problème sera formulé de même par :

$$\begin{aligned} \mathcal{M}_1 : \quad E\left[y \mid X, Z\right] \quad &= E\left[y \mid X\right] \\ \mathcal{M}_2 : \quad E\left[y \mid X, Z\right] \quad &= E\left[y \mid Z\right] \end{aligned} \tag{2.10}$$

L'exclusion de Z du modèle de régression $\mathcal{M}_1$ se fera alors sans imposer de forme fonctionnelle pour la régression et sans spécifier la loi de probabilité des variables étudiées (voir section 4.2).

2.2.2 Tests paramétriques classiques

Test de Cox :

L'un des tests les plus connus pour tester du choix entre modèles de régression linéaires non-emboîtés est dû à Cox [21] et [22], que l'on trouve explicité par Pesaran [70]. La procédure de test repose sur la différence $\widehat{L}_{12}$ entre les log-vraisemblances empiriques $\widehat{L}_1$ et $\widehat{L}_2$ des modèles $\mathcal{M}_1$ et $\mathcal{M}_2$. On examine alors la différence entre $\widehat{L}_{12}$ et sa pseudo-vraie valeur dans l'optique de $\mathcal{M}_1$. Cox obtient ainsi la statistique :

$$T_f = \frac{1}{n} \cdot \widehat{L}_{12} - E_1\left[\frac{1}{n} \cdot \widehat{L}_{12}\right]$$

où E_1 désigne l'espérance relative au modèle $\mathcal{M}_1$.

On montre alors que $\sqrt{n} \cdot T_f$ a une distribution normale centrée sous $\mathcal{M}_1$. Une procédure de test peut alors être menée en estimant la variance de T_f. Un des reproches fait à ce test est que si l'on conduit un second test sous l'hypothèse que $\mathcal{M}_2$ est vrai, les deux tests peuvent mener à des contradictions, rejetant ou acceptant simultanément les deux hypothèses concurrentes. De plus ce test ne s'applique pas si le modèle $\mathcal{M}_1$ est emboîté dans $\mathcal{M}_2$, ni si les espaces engendrés par des régresseurs X et Z sont orthogonaux (voir Pesaran).

Emboîtement artificiel :

De nombreux auteurs ont également proposé d'utiliser un modèle emboîtant les deux modèles concurrents, ainsi Atkinson [4] propose de combiner les deux modèles en un modèle général constitué d'une moyenne géométrique des densités intervenantes dans chacun des modèles. Une autre possibilité suggérée également par Atkinson consiste à réaliser une mixture des deux modèles. De cette idée provient le classique sur-modèle de régression $\mathcal{M}_c$:

$$\mathcal{M}_c \quad : \quad y = X\beta + Z\gamma + U$$

On peut alors tester l'hypothèse $\beta = 0$ (qui correspond à $\mathcal{M}_2$), puis $\gamma = 0$ (qui correspond à $\mathcal{M}_1$). Cependant, comme précédemment, les conclusions de ces tests peuvent être contradictoires. Une autre critique est qu'il n'existe pas un seul et unique sur-modèle $\mathcal{M}_c$, d'autres problèmes dûs à la colinéarité entre X et Z peuvent également affecter ces tests.

Davidson et Mac Kinnon [23] proposent en 1981 un test basé sur le modèle emboîtant suivant :

$$\mathcal{M}_c \quad : \quad y = (1-\lambda)X\beta + \lambda Z\gamma + U$$

L'idée est alors de tester la validité de l'un ou l'autre des modèles via λ. Le problème est que le modèle $\mathcal{M}_c$ n'est pas directement estimable, les paramètres β,γ et λ n'étant pas séparément identifiables. Une solution proposée est de remplacer $\mathcal{M}_c$ par un modèle $\mathcal{M}'_c$ où les paramètres d'un modèle ($\mathcal{M}_2$ par exemple) sont remplacés par un estimateur ($\widehat{\gamma}$ consistant pour $\mathcal{M}_2$) :

$$\mathcal{M}'_c \quad : \quad y = (1-\lambda)X\beta + \lambda Z\widehat{\gamma} + U \tag{2.11}$$

On teste ensuite la validité de l'autre modèle ($\mathcal{M}_1$) en testant λ.

Sur notre exemple, on teste $\mathcal{M}_1$ contre $\mathcal{M}_2$ en testant $\lambda = 0$. Si la nullité de λ est acceptée alors on validera le modèle $\mathcal{M}_1$. Un point important, relevé par Gourieroux et Monfort [38], est que la nouvelle variable $Z\widehat{\gamma}$ dépend de $(Y_i)_{i=1,..,n}$ par l'intermédiaire de $\widehat{\gamma}$ et devrait être considérée comme endogène. Cet obstacle est ignoré par Davidson et Mac Kinnon qui étudient directement

la t-statistique de λ, calculée "*comme si*" $Z\widehat{\gamma}$ était une variable exogène traditionnelle .

Dans un cadre non linéaire, deux tests reposent également sur le même principe, le **J**-test qui utilise la t-statistique pour $\lambda = 0$ dans l'estimation **J**ointe de β et λ dans (2.11), et le P-test, qui permet l'utilisation des moindres carrés linéaires dans la même situation (voir l'ouvrage de Davidson et Mac Kinnon [24]).

2.2.3 Tests d'enveloppement

Dans le cadre de modèles de régression linéaires standards, Sawa [77], nous donne l'expression des pseudo-vraies valeurs du modèle $\mathcal{M}_2$, pour cela nous noterons $\alpha = (\beta, \sigma^2)'$ les paramètres de $\mathcal{M}_1$, $\delta = (\gamma, \tau^2)'$ ceux de $\mathcal{M}_2$, X et Z les matrices de régresseurs de dimensions $(n \times p)$ et $(n \times q)$ respectivement, et où y est un vecteur d'observations de dimension $(n \times 1)$. Pour la clarté de la présentation nous supposerons que $\mathcal{M}_1$ et $\mathcal{M}_2$ sont "*strictement non emboîtés*", c'est-à-dire que la matrice $(X\ Z)$ est de rang (plein) $p+q$, la généralisation au cas où $\mathcal{M}_1$ et $\mathcal{M}_2$ sont imbriqués ne pose pas de problème majeur et est discuté dans Mizon et Richard [66].

L'estimateur du maximum de vraisemblance de $\alpha = (\beta, \sigma^2)'$ est $\widehat{\alpha} = (\widehat{\beta}, \widehat{\sigma}^2)'$ défini par :

$$\begin{aligned} \widehat{\beta} &= (X'X)^{-1} X'y \\ \widehat{\sigma}^2 &= \tfrac{1}{n} \cdot y' M_X y \end{aligned} \tag{2.12}$$

de même pour $\mathcal{M}_2$, l'estimateur de $\delta = (\gamma, \tau^2)'$ est $\widehat{\delta} = (\widehat{\gamma}, \widehat{\tau}^2)'$:

$$\begin{aligned} \widehat{\gamma} &= (Z'Z)^{-1} Z'y \\ \widehat{\tau}^2 &= \tfrac{1}{n} \cdot y' M_Z y \end{aligned} \tag{2.13}$$

Les matrices M_X et M_Z sont les matrices de projection sur les espaces orthogonaux aux espaces engendrés par X et Z respectivement. On définit également ici les matrices de projection orthogonales P_X et P_Z. Soit :

$$\begin{aligned} M_X &= I - X(X'X)^{-1}X' \quad et \quad P_X = X(X'X)^{-1}X' \\ M_Z &= I - Z(Z'Z)^{-1}Z' \qquad\qquad P_Z = Z(Z'Z)^{-1}Z' \end{aligned} \tag{2.14}$$

Pseudo-vraies valeurs

Afin d'obtenir les pseudo-vraies valeurs de $\widehat{\delta}$ sous $\mathcal{M}_1$, nous devons calculer les éléments $\Gamma(\alpha)$ et $T^2(\alpha)$ minimisant le KLIC entre $\mathcal{M}_1$ et $\mathcal{M}_2$.

Sawa [77] nous donne $\Delta(\alpha) = (\Gamma(\alpha), T^2(\alpha))'$, la pseudo-vraie valeur de $\widehat{\delta} = (\widehat{\gamma}, \widehat{\tau}^2)'$ sous $\mathcal{M}_1$:

$$\begin{aligned} \Gamma(\alpha) &= (Z'Z)^{-1} Z'X\beta \\ T^2(\alpha) &= \tfrac{1}{n} \cdot ((n-q)\sigma^2 + \beta' X' M_Z X \beta) \end{aligned}$$

Preuve :

Par définition la pseudo-vraie valeur est l'élément $\Delta(\alpha)$:

$$\Delta(\alpha) = Arg \min_{\Theta_\delta} E_\alpha \left[\log \left(\frac{L_1(\alpha)}{L_2(\delta)} \right) \right] \tag{2.15}$$

où E_α est l'espérance prise sous $\mathcal{M}_1$.

Sawa propose de séparer ce calcul à partir de l'expression de la log-vraisemblance de $\mathcal{M}_2$:

$$\log L_2(\delta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\tau^2) - \frac{1}{2\tau^2}\|y - Z\gamma\|^2$$

Si l'on différencie cette dernière expression par rapport à γ d'une part et τ^2 d'autre part, on a :

$$\begin{aligned} \tfrac{\partial \log L_2(\delta)}{\partial \gamma} &= \tfrac{1}{\tau^2} Z' (y - Z\gamma) \\ \tfrac{\partial \log L_2(\delta)}{\partial \tau^2} &= -\tfrac{n}{2\tau^2} + \tfrac{1}{2\tau^4} \|y - Z\gamma\|^2 \end{aligned}$$

La solution du problème de minimisation (2.15) est alors obtenue comme $\Delta(\alpha)$ solution de :

$$\begin{cases} E_\alpha \left[\tfrac{\partial \log L_2(\Delta)}{\partial \gamma} \right] = 0 \\ E_\alpha \left[\tfrac{\partial \log L_2(\Delta)}{\partial \tau^2} \right] = 0 \end{cases}$$

Or :

$$\begin{aligned} E_\alpha \left[\tfrac{\partial \log L_2(\delta)}{\partial \gamma} \right] &= \tfrac{1}{\tau^2} Z' (X\beta - Z\gamma) \\ E_\alpha \left[\tfrac{\partial \log L_2(\delta)}{\partial \tau^2} \right] &= -\tfrac{n}{2\tau^2} + \tfrac{1}{2\tau^4} E_\alpha \left[\|y - Z\gamma\|^2 \right] \end{aligned} \tag{2.16}$$

On peut décomposer cette dernière équation de façon à faire apparaître la variance de $\mathcal{M}_1$:

$$\begin{aligned} E_\alpha \left[\tfrac{\partial \log L_2(\delta)}{\partial \tau^2} \right] &= -\tfrac{n}{2\tau^2} + \tfrac{1}{2\tau^4} E_\alpha \left[\|y - Z\gamma\|^2 \right] \\ &= -\tfrac{n}{2\tau^2} + \tfrac{1}{2\tau^4} E_\alpha \left[\|y - X\beta\|^2 + \|X\beta - Z\gamma\|^2 \right] \\ &= -\tfrac{n}{2\tau^2} + \tfrac{1}{2\tau^4} \left[(n-q)\sigma^2 + \|X\beta - Z\gamma\|^2 \right] \end{aligned} \tag{2.17}$$

On obtient $\Gamma(\alpha)$ et $T^2(\alpha)$ en déterminant les éléments γ et τ^2 réalisant l'égalité à zéro des expressions (2.16) et (2.17) respectivement. □

Remarque :

Une interprétation géométrique de ce résultat est que $Z \cdot \Gamma(\alpha)$ est la projection de l'espérance $(X\beta)$ de y sous $\mathcal{M}_1$ sur l'espace engendré par Z. En effet,

$$Z \cdot \Gamma(\alpha) = Z\left(Z'Z\right)^{-1} Z' \cdot X\beta = P_Z \cdot X\beta$$

tandis que $nT^2(\alpha)$ est la somme des variances des y_i à laquelle s'ajoute la norme euclidienne de la distance entre les espérance de y sous $\mathcal{M}_1$ $(X\beta)$ et celle sous $\mathcal{M}_2$ $(Z\gamma)$.

Il est aisé de montrer que :

Lemme 1 *: Le comportement asymptotique des estimateurs du maximum de vraisemblance sous $\mathcal{M}_1$ est :*

i) $E_\alpha(\widehat{\gamma}) = \Gamma(\alpha)$

ii) $Lim_{n\to\infty}\ E_\alpha(\widehat{\tau}^2 - T^2(\alpha)) = 0$

Ce lemme dû à Sawa [77], nous permet de vérifier que l'espérance de l'estimateur du maximum de vraisemblance sous une mauvaise spécification, donne la pseudo-vraie valeur. Celle-ci minimise la distance, au sens de Kullback-Leibler, entre le modèle de référence, $\mathcal{M}_1$, et le modèle par rapport auquel est calculé cet estimateur, $\mathcal{M}_2$.

Statistique de test

La statistique $\phi = \delta - \Delta(\alpha)$ définissant la différence d'enveloppement s'écrit alors comme le vecteur :

$$\phi = \begin{pmatrix} \gamma - \Gamma(\alpha) \\ \\ \tau^2 - T^2(\alpha) \end{pmatrix} = \begin{pmatrix} \left(Z'Z\right)^{-1} Zy - \left(Z'Z\right)^{-1} Z'X\beta \\ \\ \frac{1}{n} y'M_Z y - \frac{1}{n}\left((n-q)\sigma^2 + \beta'X'M_Z X\beta\right) \end{pmatrix}$$

Cette pseudo-vraie valeur est estimée par $\widehat{\phi}$:

$$\widehat{\phi} = \begin{pmatrix} \widehat{\gamma} - \widehat{\Gamma}(\alpha) \\ \\ \widehat{\tau}^2 - \widehat{T}^2(\alpha) \end{pmatrix} = \begin{pmatrix} \left(Z'Z\right)^{-1} Z'\, y - \left(Z'Z\right)^{-1} Z'X\widehat{\beta} \\ \\ \frac{1}{n} \cdot y'M_Z y - \frac{1}{n}\left((n-q)\widehat{\sigma}^2 + \widehat{\beta}'X'M_Z X\widehat{\beta}\right) \end{pmatrix}$$

soit encore :

$$\widehat{\phi} = \begin{pmatrix} (Z'Z)^{-1} Z'(y - X\widehat{\beta}) \\ \\ \frac{1}{n} \cdot y' \left[M_Z - (n-q) M_X \right] y - \frac{1}{n} \left[\widehat{\beta}' X' M_Z X \widehat{\beta} \right] \end{pmatrix}$$

L'utilisation des formules (2.12) et (2.13) permet de simplifier l'écriture de cette différence, où l'on remarque que :

- $X\widehat{\beta}$ est la projection de y sur l'espace engendré par X et donc $X\widehat{\beta} = P_X y$
- de fait, $y - X\widehat{\beta} = M_X y$, et on a :

$$\widehat{\phi} = \begin{pmatrix} (Z'Z)^{-1} Z' M_X y \\ \\ \frac{1}{n} y' \left(M_Z - (n-q)' M_X - P_X M_Z P_X \right) y \end{pmatrix} \tag{2.18}$$

La première coordonnée de $\widehat{\phi}$ s'exprime donc comme étant une expression linéaire en y, la deuxième est une forme quadratique en y.

La variance asymptotique $V_\alpha(\widehat{\phi})$ de la statistique $\sqrt{n} \cdot \widehat{\phi}$ est :

$$V_\alpha(\widehat{\phi}) = \begin{pmatrix} n\sigma^2 (Z'Z)^{-1} Z' M_X Z (Z'Z)^{-1} & -2\sigma^2 (Z'Z)^{-1} Z' P_X M_Z X\beta \\ \\ -2\sigma^2 \beta' X' M_Z P_X Z (Z'Z)^{-1} & \frac{4\sigma^2}{n} \cdot \beta' X' M_Z M_X M_Z X\beta \end{pmatrix} \tag{2.19}$$

On peut également l'écrire sous la forme :

$$V_\alpha(\widehat{\phi}) = n\sigma^2 \cdot Q \ (Z'Z)^{-1} Z' M_X Z (Z'Z)^{-1} Q'$$

où $Q' = \begin{pmatrix} I_q & -\frac{2}{n} Z' X \beta \end{pmatrix}$.

Une preuve de ce résultat dû à Mizon et Richard est rappelée en annexe.

Ces premiers résultats nous permettent d'obtenir différents tests d'enveloppement suivant le paramètre d'intérêt.

Tests de Wald

Trois tests d'enveloppement de Wald sont proposés ici selon que l'on s'intéresse au paramètre "*complet*" $\delta = (\gamma, \tau^2)'$ ou selon que l'on envisage l'enveloppement sur la première (ou deuxième) coordonnée pour ne retenir qu'un test d'enveloppement sur $\widehat{\gamma}$ (ou $\widehat{\tau}^2$).

- Le test d'enveloppement complet est défini par :

$$\eta(\widehat{\delta}) = n \cdot \widehat{\phi}' \cdot V_{\widehat{\alpha}}(\widehat{\phi})^{+} \cdot \widehat{\phi}$$

où : $V_{\widehat{\alpha}}(\widehat{\phi})$ est un estimateur de $V_{\alpha}(\widehat{\phi})$ obtenu en estimant les paramètres de α,

et $V_{\widehat{\alpha}}(\widehat{\phi})^{+}$ désigne un inverse généralisé de cette matrice.

Cette statistique est asymptotiquement distribuée sous $\mathcal{M}_1$, suivant une loi χ^2 à q degrés de liberté[3].

- Deux statistiques "*marginales*" peuvent être extraites de ces expressions, $\eta(\widehat{\gamma})$ est ainsi basé sur le coefficient de régression $\widehat{\gamma}$ tandis que $\eta(\widehat{\tau}^2)$ porte sur l'estimateur de la variance $\widehat{\tau}^2$. On obtient ainsi :

$$\eta(\widehat{\gamma}) = \left(\widehat{\gamma} - \widehat{\Gamma}(\widehat{\beta})\right)' V_{\widehat{\alpha}}(\widehat{\gamma})^{+} \left(\widehat{\gamma} - \widehat{\Gamma}(\widehat{\beta})\right)$$

où : $V_{\widehat{\alpha}}(\widehat{\gamma}) = n\widehat{\sigma}^2 \left(Z'Z\right)^{-1} Z' M_X' Z \left(Z'Z\right)^{-1}$.

Ce qui nous donne :

$$\eta(\widehat{\gamma}) = \frac{1}{\widehat{\sigma}^2} \cdot y' M_X Z \left(Z' M_X Z\right)^{-1} Z' M_X y$$

Sous $\mathcal{M}_1$, cette statistique est, elle aussi, asymptotiquement distribuée suivant une loi χ^2 à q degrés de liberté[4].

- La troisième statistique d'intérêt, porte sur la différence des variances estimées de $\mathcal{M}_2$

$$\eta(\widehat{\tau}^2) = \left(\widehat{\tau}^2 - \widehat{T}^2(\widehat{\alpha})\right)' V_{\widehat{\alpha}}(\widehat{\tau}^2)^{+} \left(\widehat{\tau}^2 - \widehat{T}^2(\widehat{\alpha})\right)$$

où : $V_{\widehat{\alpha}}(\widehat{\tau}^2) = \frac{4\sigma^2}{n} \cdot \beta' X' M_Z M_X M_Z X \beta$.

Cette statistique est asymptotiquement distribuée sous $\mathcal{M}_1$, suivant une loi $\chi^2(1)$.

Il est intéressant de noter que ces statistiques sont en relation entre elles et avec des statistiques classiques.

---

[3]Le nombre de degrés de liberté est en fait le rang de $V_{\alpha}(\widehat{\phi})$, qui correspond au nombre de variables propres au modèle $\mathcal{M}_2$, en supposant les modèles strictement non emboîtés, le nombre de degrés de liberté est donc q.

[4]Le rang de $V_{\alpha}(\widehat{\gamma})$ est le même que le rang de $V_{\alpha}(\widehat{\phi})$, c'est le rang de la matrice $\left(Z'Z\right)^{-1} Z' M_X Z \left(Z'Z\right)^{-1}$.

- La F-statistique relative à l'hypothèse $\gamma_* = 0$ dans la régression emboîtante M_c :

$$\mathcal{M}_c \quad : \quad y = X\beta + Z\gamma_* + U$$

donne :

$$F_c = \frac{n-p-q}{nq} \cdot \frac{\widehat{\gamma}_*' Z' M_X Z \widehat{\gamma}_*}{\widehat{\tau}_*^2}$$

où : $\widehat{\gamma}_*' = (Z' M_X Z)^{-1} Z' M_X \; y$ et $n\widehat{\tau}_*^2 = y' M_X y - \widehat{\gamma}_*' Z' M_X Z \widehat{\gamma}_*$

un rapide calcul nous permet d'exprimer cette statistique comme étant :

$$qF_c = \frac{n-p-q}{n} \cdot \eta(\widehat{\delta}) \left[1 - \frac{1}{n}\eta(\widehat{\delta})\right]^{-1}$$

les statistiques $\eta(\widehat{\delta})$ et qF_c sont par conséquent asymptotiquement équivalentes sous $\mathcal{M}_1$.

- Mizon et Richard [66] montrent que la statistique de test $\eta(\widehat{\gamma})$ est équivalente asymptotiquement à la statistique "*complète*" $\eta(\widehat{\delta})$.

En effet on a l'équivalence asymptotique sous $\mathcal{M}_1$ $\left(\overset{\mathcal{M}_1}{\sim}\right)$:

$$\sqrt{n} \cdot \left(\widehat{\tau}^2 - \widehat{T}^2(\widehat{\alpha})\right) \overset{\mathcal{M}_1}{\sim} -2\beta' \left(\frac{X'Z}{n}\right) \sqrt{n} \cdot \left(\widehat{\gamma} - \widehat{\Gamma}(\widehat{\beta})\right)$$

En outre, la même équivalence entre statistique complète et statistique sur $\widehat{\gamma}$ se retrouve lorsque l'on examine l'enveloppement de $\mathcal{M}_c$ par $\mathcal{M}_1$.

- Hendry et Richard [54] montrent également que la statistique complète relative à l'enveloppement de $\mathcal{M}_c$ par $\mathcal{M}_1$ est équivalente asymptotiquement à la statistique complète d'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$.

Ce dernier point permet notamment de réconcilier les approches emboîtées et non-emboîtées, qui sont ainsi équivalentes dans cette approche.

Remarque :

Les statistiques $\eta(\widehat{\gamma})$ et $\eta(\widehat{\tau}^2)$ peuvent être retrouvées par l'utilisation de la notion d'enveloppement via une fonction $\mathcal{B}$ définie section 2.1.3. Si l'on utilise les fonctions $\mathcal{B}_1$ et $\mathcal{B}_2$ définies ci - dessous :

$$\begin{array}{rcccc} \mathcal{B}_1 & : & \Re^q \times \Re^+ & \longrightarrow & \Re^q \\ & & \begin{pmatrix} \widehat{\gamma} \\ \widehat{\tau}^2 \end{pmatrix} & \longrightarrow & \widehat{\gamma} \end{array}$$

et

$$\begin{array}{rcccc} \mathcal{B}_1 & : & \Re^q \times \Re^+ & \longrightarrow & \Re^+ \\ & & \begin{pmatrix} \widehat{\gamma} \\ \widehat{\tau}^2 \end{pmatrix} & \longrightarrow & \widehat{\tau}^2 \end{array}$$

nous retrouvons les statistiques de tests $\eta(\widehat{\gamma})$ et $\eta(\widehat{\tau}^2)$ à partir des formules générales 2.3 et 2.4 données dans la section 2.1.3.

2.3 Conclusion

Les tests d'enveloppement présentés dans ce chapitre reposent sur la définition de l'enveloppement approché et se fondent sur l'étude du défaut d'enveloppement, constitué de la différence entre un estimateur des paramètres du modèle $\mathcal{M}_2$ et un estimateur de la pseudo-vraie valeur sous $\mathcal{M}_1$. Cette différence est examinée de manière globale ou partielle selon que l'on intègre l'ensemble des paramètres des modèles, ou une partie seulement. Une classe de tests de Wald examinant le défaut d'enveloppement par l'intermédiaire d'une fonction déterministe ou non, est développée et permet une généralisation des tests existants. Cette approche regroupe sous une même présentation une vaste collection de tests d'hypothèses emboîtées et non-emboîtées, et permet de retrouver les tests classiques comme des cas particuliers.

Ces tests doivent cependant être considérés comme des test visant à comparer les forces et faiblesses des modèles en présence et non comme des procédures de validation ou de sélection. Notre approche est ainsi directionnelle considérant le modèle $\mathcal{M}_1$ comme modèle d'intérêt que l'on cherche à valider par ses capacité à incorporer les résultats de modèles secondaires qui ne sont que les instruments de cette validation.

Nous présenterons chapitre 4 différents tests permettant d'examiner la validation d'un modèle de régression par l'enveloppement d'un autre modèle de régression dans un contexte non-paramétrique. Il nous faut auparavant définir les estimateurs non-paramétriques qui interviendront dans la définition de ces modèles.

2.4 Annexe au chapitre 2

Autre possibilité d'énoncer le théorème 2.3 :

Les dérivées premières et secondes de $B(Y_n, \gamma)$ jouant un grand rôle dans cette distribution, nous devons auparavant introduire quelques notations simplifiant l'écriture (voir Mizon [65]) :

- $\overline{B}(\beta, \gamma) = p\lim_{\mathcal{M}_1} B^1(Y_n, \gamma)$ où $B^1(Y_n, \gamma) = \left(\frac{\partial B(Y_n,\gamma)}{\partial \gamma'}\right)$
- $\phi^* = B(Y_n, \Gamma(\beta)) - E_{\mathcal{M}_1}\left[B\left(Y_n, \Gamma(\beta)\right)\right] - \phi_\beta\left(\beta - \widehat{\beta}\right)$

  avec : $\phi_\beta = \lim_{n\to\infty} \frac{1}{n} E_\beta\left(B(Y_n, \Gamma(\beta)) \cdot \frac{\partial L_1}{\partial \beta}\right)$

Théorème 2.5 *Sous les hypothèses du théorème 2.1, et sous les hypothèses de régularité suivantes :*

- *La fonction $B\left(Y_n, \gamma\right)$ est de classe C^1 par rapport à γ sur un voisinage V de $\Gamma(\beta)$*
- *La matrice $\overline{B}$ est finie dans le voisinage V et de rang r*
- *Les matrices $p\lim_{\mathcal{M}_1}\left(\frac{\partial B^1(Y_n,\gamma)}{\partial \gamma_i}\right)_\gamma$ sont finies pour $\gamma \in V$*

  On a,

$$\sqrt{n}\widetilde{\phi} = \sqrt{n}\left(\widetilde{\gamma} - E_{\mathcal{M}_1}\widetilde{\gamma}\right) = \sqrt{n}\phi^* + \overline{B}(\beta, \Gamma(\beta)) \cdot \sqrt{n}\widehat{\phi} + o_p(1)$$

□

Calcul de la variance $V_\alpha(\widehat{\phi})$ donné par la formule (2.19) :

La formule (2.18) nous donne une expression de $\widehat{\phi}$ comme étant une expression linéaire en y pour la première composante, la deuxième étant une forme quadratique en y. Soit encore :

$$\widehat{\phi} = \begin{pmatrix} \widehat{\gamma} - \widehat{\Gamma}(\alpha) \\ \\ \widehat{\tau}^2 - \widehat{T}^2(\alpha) \end{pmatrix} = \begin{pmatrix} Ay \\ \\ \frac{1}{n} \cdot y'By \end{pmatrix}$$

où $A = \left(Z'Z\right)^{-1} Z'M_X$ et $B = M_Z - (n-q)' M_X - P_X M_Z P_X$.

Sous $\mathcal{M}_1$, $y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$ et donc :

$$Var\left(\widehat{\gamma} - \widehat{\Gamma}(\alpha)\right) = \sigma^2 AA' = \sigma^2 \left(Z'Z\right)^{-1} Z' M_X Z \left(Z'Z\right)^{-1}$$

On a également :

$$Cov\left(\widehat{\gamma} - \widehat{\Gamma}(\alpha), \widehat{\tau}^2 - \widehat{T}^2(\alpha)\right) \quad = \quad \tfrac{2\sigma^2}{n} ABX\beta$$

on remarque que $BX\beta = M_X M_Z X\beta$, ce qui nous donne :

$$Cov\left(\widehat{\gamma} - \widehat{\Gamma}(\alpha), \widehat{\tau}^2 - \widehat{T}^2(\alpha)\right) \quad = \quad \tfrac{2\sigma^2}{n} \left(Z'Z\right)^{-1} Z' M_X M_Z X\beta$$

et,

$$Var\left(\widehat{\tau}^2 - \widehat{T}^2(\alpha)\right) = \frac{2\sigma^4}{n^2} Tr(B^2) + \frac{4\sigma^2}{n^2} \beta' X' M_Z M_X M_Z X\beta$$

De cette dernière expression, seul le deuxième terme apparaît dans l'expression (2.19), le premier terme étant négligeable devant ce terme (la trace $Tr(B^2)$ est en effet un $O_p(\frac{1}{n^2})$ voir Mizon et Richard [66]). □

[a4,12pt]freport

Chapter 3

Estimation non-paramétrique de la régression

Ce chapitre se veut une introduction aux estimateurs non-paramétriques de la régression. Nous présenterons également les principales propriétés que nous utiliserons dans notre étude non-paramétrique de l'enveloppement. Après avoir présenté succinctement différents estimateurs fonctionnels de la régression nous détaillerons plus particulièrement les propriétés relatives à la méthode du noyau. Nous nous préoccuperons enfin du problème de sélection de la fenêtre. Ce chapitre ne présente toutefois aucun apport statistique nouveau.

3.1 Introduction

"On dit qu'un problème d'estimation est non-paramétrique lorsqu'il ne peut pas se ramener au problème de l'estimation d'un élément d'un espace vectoriel de dimension finie"

Gerard Collomb (1976)

D'après G. Collomb, l'estimation non-paramétrique se présente comme une "*non-définition*", rejetant l'estimation d'un paramètre sans que ne soit explicitement exposé l'objet à estimer.

Dans notre approche non-paramétrique l'objet d'intérêt est une fonction tout-à-fait générale, appartenant à un espace fonctionnel (ce qui n'exclut cependant pas tout paramètre de l'estimation).

Dans le cadre de ce travail, et afin de clarifier notre propos, nous entendrons par "non-paramétrique" l'estimation du modèle de régression :

$$Y = f(x) + u$$

où $E[u \mid X] = 0$, $p.s.$

dans laquelle, ni la forme de la fonction de régression, ni la distribution des résidus ne seront spécifiés.

Ceci est la double négation d'un modèle paramétrique où, par exemple, la forme linéaire est imposée et où la distribution des résidus est spécifiée.

Une classification précise entre estimation paramétrique, non-paramétrique, semi-paramétrique et semi-non-paramétrique nous a été présentée par M. Delecroix et est reproduite dans le travail de Pascal Lavergne [58]. Cette classification se base sur l'objet d'intérêt de l'estimation, et nous ne la détaillerons pas davantage, laissant le lecteur intéressé se reporter à ces références.

L'accent sera mis dans ce chapitre sur l'importance des choix arbitraires intervenant dans l'estimation non-paramétrique, et en premier lieu, sur le choix du paramètre déterminant le degré de "*douceur*" de l'estimateur non-paramétrique. En effet, "*non-paramétrique*" ne signifie pas absence de paramètre, bien au contraire et un paramètre de lissage interviendra de manière cruciale dans l'estimation. A travers les résultats asymptotiques et les exemples d'estimateurs classiques proposés, nous essayerons de relever l'aspect d'arbitrage que revêt ce paramètre entre "*douceur*" et "*variabilité*" des estimateurs. La sélection de ce paramètre dans le cadre de la méthode du noyau de convolution sera étudiée afin de mieux cerner l'impact de ce choix sur l'estimateur. Le critère de la validation croisée sera retenu pour la suite de notre travail et nous tenterons de motiver ce choix en relevant le caractère *objectif* de ce critère face à l'*arbitraire* des choix *ad hoc*.

Notre présentation des estimateurs de la régression classiques s'inspire du cours de M. Delecroix, et de la revue bibliographique de Collomb [19]. Nous retrouvons ainsi la modélisation par δ-suites dues à Walter et Blum et portées à notre connaissance par B. Portier et P. Ango-Nzé [71], que l'on trouve également dans Rao ([72] pp.135-143).

Le problème de l'inexistence d'un estimateur sans biais de la régression sur un échantillon fini montré par Collomb ([17] pp.12-15) sera contourné par l'utilisation systématique d'une optique asymptotique. Le biais sera toutefois analysé et des procédures pour "tuer" ce biais seront exposées. Il s'agira principalement de contraintes sur le paramètre de lissage.

Afin d'assurer l'existence de $f(x) = E[Y \mid X = x]$, nous supposerons que $E[|Y|] < \infty$. La fonction f n'étant définie sur $\Re^p$ qu'à une équivalence près, nous supposerons également qu'il en existe une version continue f. Par convention, on posera $f(x) = 0$ si $\varphi(x) = 0$.

3.2 Définition des estimateurs

L'estimation non-paramétrique de la régression repose sur l'idée intuitive que l'estimateur $\widehat{f}(\cdot)$ en un point x doit être "*proche*" de Y_i si x est "*proche*" de X_i. La même propriété se répétant sur l'ensemble des observations, les estimateurs non-paramétriques de la régression s'écriront donc comme des moyennes pondérées des Y_i, la pondération prenant en compte l'éloignement de X_i au point considéré. Par souci de clarté nous nous restreindrons momentanément au cas particulier univarié ($p = 1$).

La forme générale d'un estimateur non-paramétrique de la régression, tel que nous venons de le présenter, sera donc :

$$\widehat{f}(x) = \sum_{i=1}^{n} Y_i \cdot W_m(X_i, x)$$

De manière à obtenir une pondération de somme totale unitaire, on posera :

$$W_m(X_i, x) = \frac{w_m(X_i, x)}{\sum_{i=1}^{n} w_m(X_i, x)}$$

Suivant le type de pondération utilisé, nous obtiendrons différents types d'estimateurs, chacun dépendant d'un paramètre dont le choix permet de déterminer la "*douceur*" de l'estimateur :

- L'estimateur des k points les plus proches :

$$w_k(X_i, x) = \mathbf{1}_{[les\ k\ X_j\ les\ plus\ proches\ de\ x]}\ (X_i)$$

  le paramètre intervenant dans cette définition est le nombre k de points considérés comme pertinents pour estimer f en x, les autres points ne participant pas au calcul de $\widehat{f}(x)$.

- L'estimateur de la fenêtre mobile (ou régressogramme) est obtenu en posant :

$$w_h(X_i, x) = \mathbf{1}_{[x-h\ ,\ x+h]}(X_i)$$

  Ici, le paramètre est la largeur $2h$ de la fenêtre[1] dans laquelle sont retenues les observations sur lesquelles porte la moyenne des Y_i.

  Ces deux estimateurs, s'ils permettent l'estimation de la régression en tous points, ne sont pas continus et présentent des sauts dûs à la fonction indicatrice intervenant dans leur définition.

[1] C'est à cette technique que le paramètre de lissage doit le terme usuel de fenêtre.

- L'estimateur du noyau de convolution

$$w_h(X_i, x) = K\left(\frac{X_i - x}{h}\right)$$

où K est une fonction continue prenant en compte l'ensemble des points de l'échantillon, dont la valeur diminue avec l'éloignement entre x et X_i, et où h est un réel positif permettant de *relativiser* l'éloignement de x à X_i. Cet estimateur sera étudié plus en détail section 3.3.

- L'estimateur du noyau de convolution récursif

$$w_h(X_i, x) = K\left(\frac{X_i - x}{h_i}\right)$$

est une variante de cet estimateur introduit par Devroye et Wagner (Voir Härdle [46]), pour lequel h est remplacé par la suite $(h_i)_{i=1,\ldots,n}$, h_i variant avec le point X_i considéré.

- L'estimateur des fonctions orthogonales

$$w_M(X_i, x) = \sum_{j=1}^{M} e_j(X_i) \cdot e_j(x)$$

La suite de fonctions $(e_j(\cdot))_{j \in Z}$ constitue une base orthogonale de l'espace Hilbertien $L^2(\Re)$ supposé contenir la fonction de régression f. L'entier M représente le nombre d'éléments de la base intervenant dans cet estimateur.

- L'estimateur à ondelettes orthogonales

$$w_p(X_i, x) = \sum_{j \in Z} 2^p \Phi(2^p X_i - j) \cdot \Phi(2^p x - j)$$

où $\Phi(\cdot) \in L^2(\Re)$ est issue d'une analyse multirésolution, qui est une décomposition de l'espace $L^2(\Re)$ en une suite croissante d'espaces vectoriels fermés V_j. On établit que pour tout entier relatif j, la suite de fonctions $\left(2^{\frac{j}{2}} \Phi\left(2^j \cdot -k\right)\right)_{k \in \mathbf{Z}}$ constitue une base orthonormée de V_j. Le paramètre p (ou plutôt 2^p) détermine la "finesse" de cette décomposition., (voir Portier [71], ou Gasquet et Witomsky [34]).

D'autres estimateurs peuvent s'écrire sous cette forme avec une pondération non unitaire et sont proposés par Ullah et Vinod [85].

- L'estimateur de Mack et Müller [62]

$$W_h(X_i, x) = \frac{1}{nh} \frac{1}{\widehat{\varphi}(X_i)} \cdot K\left(\frac{X_i - x}{h}\right)$$

où $\widehat{\varphi}(X_i) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right)$ est l'estimateur du noyau de la densité φ au point X_i.

Cet estimateur est l'estimateur du noyau pour lequel le dénominateur est $\widehat{\varphi}(X_i)$ au lieu de $\widehat{\varphi}(x)$. L'avantage de cet estimateur réside dans le calcul de l'estimateur de la dérivée de $f(x)$ puisque seul le numérateur est alors à calculer.

- L'estimateur de Gasser et Müller [35]

$$W_h(X_i, x) = \frac{1}{h} \int_{A_i} K\left(\frac{u - x}{h}\right) du$$

où $A_i = (S_{i-1}, S_{i+1})$, et $S_i = (X_i + X_{i+1})/2$ est le milieu des points X_i et X_{i+1}. Cet estimateur est développé dans le cadre des régresseurs fixes.

D'autres techniques d'estimation non-paramétrique existent néanmoins et ne sont pas issues de la même logique. Certains estimateurs sont ainsi définis non pas comme une somme pondérée, mais comme minimisant un critère sur un ensemble de fonctions.

- On peut ainsi citer les "*fonctions splines cubiques*" déterminées comme les fonctions deux fois différentiables minimisant :

$$S_\lambda = \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \lambda \int (f''(x))^2 \, dx$$

sur un intervalle compact. La solution unique de cette minimisation est $\widehat{f}_\lambda$. On montre que $\widehat{f}_\lambda$ est un polynôme cubique entre deux observations successives, voir Eubank [30].

Le paramètre λ sert en fait d'arbitrage entre la *"fidélité aux données"*, $\left(\sum_i (Y_i - f(X_i))^2\right)$ et la "*douceur de la courbe*",$\left(\int (f''(x))^2 \, dx\right)$.

Silverman [**?**], remarque toutefois que l'on peut encore approximer $\widehat{f}_\lambda$ par une somme pondérée de Y_i avec la pondération suivante :

$$W_\lambda(x, X_i) = \frac{1}{\varphi(x)} \frac{1}{h_\lambda(x)} \cdot K\left(\frac{x - X_i}{h_\lambda(x)}\right)$$

  - $h_\lambda(x)$ est une fenêtre locale dépendant de λ et de la densité φ des X_i
    K est une fonction telle que $\lim_{|u| \to \infty} K(u) \longrightarrow 0$

- Une autre classe d'estimateurs définis dans la même logique de minimisation est celle des M-estimateurs à noyau solutions d'un problème d'optimisation du type :

$$\widehat{f_n}(x) = Arg \min_f \sum_{i=1}^{n} \frac{1}{h} \cdot K\left(\frac{S(W_i, \alpha_i) - x}{h}\right) \cdot \psi\left(W_i, \alpha_n, x, f\right) \qquad (3.1)$$

où S et ψ sont des fonctions connues définissant la variable conditionnante $S(W_i, \alpha_i)$ et la fonction d'objectif paramétrique respectivement. Les $(W_i)_{i=1,\ldots,n}$ sont les observations et α_i tend vers une limite α_∞ lorsque n tend vers l'infini. Cette formulation générale permet également de retrouver certains des estimateurs que nous venons de citer dans une optique de type "*condition de moment*"[2]. (voir Gourieroux, Montfort et Tenreiro[41], ou Härdle[46]).

Remarque :

- Quelle que soit la méthode d'estimation proposée, le problème de "*douceur de l'estimateur*" évoqué ci-dessus, se posera par l'intermédiaire du paramètre de lissage. Son choix permettra d'arbitrer entre "*variabilité*" et "*lissage*" ou, d'une manière plus formelle, entre "*variance*" et "*biais*" comme nous le verrons section 3.5.

- Pour chacune des méthodes envisagées ci-dessus, ce paramètre est, en réalité, dépendant de la taille de l'échantillon, voire de l'échantillon lui-même si l'on souhaite le sélectionner convenablement. Les estimateurs proposés ci-dessus seront donc retrouvés dans la littérature munis d'une suite de paramètres, $h(n), M(n), etc \ldots$. Les propriétés de convergence de ces estimateurs seront conditionnées par des hypothèses sur la vitesse de convergence (ou de divergence) de ces suites.

3.3 Estimateur du noyau de convolution

Définition 3.1 *(Noyau de Parzen-Rosenblatt)* **:**

Un noyau K est une application de $\Re^p$ dans $\Re$, bornée, intégrable pour la mesure de Lebesgue, d'intégrale unitaire. Un noyau de Parzen-Rosenblatt vérifie de plus :

$$\lim_{\|x\| \to \infty} \|x\|^p K(x) = 0$$

où $\|\cdot\|$ désigne la norme de $\Re^p$.

[2]L'estimateur à noyau de la régression est ainsi solution de (3.1) lorsque :

$$\psi\left(W_i, \alpha_n, x, f\right) = \left(Y_i - f(x)\right)^2$$

Un exemple de noyau de Parzen-Rosenblatt est la densité normale standard qui vérifie cette condition. Nous utiliserons ces noyaux dans la suite de ce travail.

On définit également des classes de noyaux correspondants à des propriétés de régularité particulières.

Définition 3.2 *(Noyau d'ordre m) :*

Le noyau K appartient à la classe $\mathcal{K}_m(\Re^p)$ des noyaux d'ordre m si :

$$\int_{\Re^p} \prod_{i=1}^{p} x_i^{a_i} K(x_1, x_2, \ldots, x_p) dx_1 \cdots dx_p = \begin{cases} 1 & si \quad a_i = 0 \ , \ \forall\, i = 1, \ldots, p \\ 0 & si \quad 0 < \sum_{i=1}^{p} a_i < m \end{cases}$$

$$et \qquad \int_{\Re^p} |x_i|^m \left|K(x_1, x_2, \ldots, x_p)\right| dx_1 \cdots dx_p < \infty \quad , \ \forall x \in \Re^p$$

Cette propriété est standard en statistique non-paramétrique et est couramment utilisée comme hypothèse de régularité pour les noyaux dans les théorèmes de convergence asymptotique.

Il est à noter que pour $m \geq 3$, les noyaux de $\mathcal{K}_m(\Re^p)$ ne sont plus des densités, et pourront prendre des valeurs négatives sur certains intervalles.

Définition 3.3 *(Estimateur du noyau de la régression) :*

Soit $(X_i, Y_i)_{i=1,\cdots,n}$ n observations d'un couple (X, Y) de variables aléatoires définies sur l'espace réel mesuré $(\Re^p \times \Re, \mathcal{B}_{\Re^p+1}, \lambda)$. L'estimateur du noyau de convolution de la régression $f(x) = E[Y \mid X = x]$ associé au noyau K et à la fenêtre h_n, un nombre réel dépendant de n, est défini par :

$$\widehat{f_n}(x) = \frac{\frac{1}{nh_n^p} \sum_{i=1}^{n} \ Y_i K\left(\frac{X_i - x}{h_n}\right)}{\frac{1}{nh_n^p} \sum_{i=1}^{n} \ K\left(\frac{X_i - x}{h_n}\right)} \qquad \forall x \in \Re^p \tag{3.2}$$

avec la convention $\widehat{f_n}(x) = 0$ si le dénominateur $\frac{1}{nh_n^p} \sum_{i=1}^{n} \ K\left(\frac{X_i - x}{h}\right) = 0$.

Cette formulation ayant été introduite simultanément par Nadaraya [67] et Watson [89] en 1964, cet estimateur est également appelé estimateur de Nadaraya -Watson.

Le dénominateur de l'expression (3.2) est un estimateur de la densité marginale $\varphi(x)$, tandis que le numérateur constitue un estimateur de $\Phi(x) = \int y\varphi(x, y)dy$. Nous pouvons donc écrire $\widehat{f_n}(x)$ sous la forme :

$$\widehat{f_n}(x) = \frac{\widehat{\Phi_n}(x)}{\widehat{\varphi_n}(x)}$$

Si, en particulier, K est une densité de probabilité alors l'estimateur $\widehat{\varphi_n}(x)$ de $\varphi(x)$ est donné par la densité de la somme de deux variables aléatoires :

- l'une suivant la densité empirique des X_i, $\mu_n = \frac{1}{n}\sum_i \delta_{X_i}$
- l'autre suivant la distribution de densité $K_n(\cdot) = \frac{1}{h_n^p} K\left(\frac{\cdot}{h_n}\right)$

La convolution ainsi réalisée suit une distribution de densité :

$$\widehat{\varphi_n}(x) = (K_n * \mu_n)(x) = \frac{1}{nh_n^p}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)$$

d'où l'estimateur tire son nom "d'estimateur du noyau de convolution".

3.4 Propriétés

Le premier résultat de convergence est dû à Bochner [11], dont le lemme est à la base des principaux théorèmes de convergence. Nous tirons son énoncé de l'ouvrage de Bosq et Lecoutre [12] (p. 61).

Lemme 2 *(de Bochner) :*

i) Soit K un noyau de Parzen-Rosenblatt et g une fonction de $\mathbf{L}^1$.
Alors, en tout point x, où g est continue :

$$\lim_{h_n \to 0} (K_n * g)(x) = g(x)$$

ii) Soit K un noyau quelconque et g une fonction de $\mathbf{L}^1$ uniformément continue, alors

$$\lim_{h_n \to 0} \sup_x |(K_n * g)(x) - g(x)| = 0$$

L'interprétation de ce lemme est que lorsque la fenêtre h est "petite", la convolution d'une fonction de $\mathbf{L}^1$ avec K_n perturbe peu cette fonction.

3.4.1 Propriétés de convergence ponctuelle de l'estimateur $\widehat{f_n}$:

Quelques hypothèses standard permettant l'application du lemme de Bochner sont présentées ici.

Hypothèse 3.1 *:*

Les observations $(X_i, Y_i)_{i=1,\dots,n}$ *sont des observations indépendantes du couple de variables aléatoires* (X, Y) *de* $\Re^p \times \Re$.

Hypothèse 3.2 *:*

Le noyau $K(\cdot)$ *est de Parzen-Rosenblatt*

Hypothèse 3.3 *:*

La fenêtre h_n *vérifie* $\lim_{n\to\infty} h_n = 0 \quad et \quad \lim_{n\to\infty} nh_n^p = \infty$

Cette dernière hypothèse sur les fenêtres est la traduction d'un arbitrage entre variabilité et douceur de l'estimateur déjà évoqué.

En imaginant visuellement deux cas limites nous voyons que :

- si $h_n \to \infty$ alors $K\left(\frac{X_i - x}{h_n}\right) \to k = K(0)$,$\forall x$ et $\widehat{f_n}(x) \to k \cdot \frac{1}{n}\sum_i Y_i = k \cdot \overline{Y}$

  L'estimateur dégénère en une fonction d'une douceur extrême puisque constante, mais sans estimer réellement $f(x)$.

- si $h_n \to 0$ sans restriction, $K\left(\frac{X_i - x}{h_n}\right) \to \begin{cases} 1 \; si \; x = X_i \\ 0 \; Sinon \end{cases}$

  L'estimateur devient alors extrêmement sensible à l'éloignement de x à X_i et tend vers une fonction discontinue passant par tous les points (X_i, Y_i).

L'hypothèse 3.3 nous impose un juste milieu, nécessaire à la convergence de l'estimateur $\widehat{f_n}$. Dans la suite de ce travail nous supposerons ces hypothèses vérifiées, et nous ne mentionnerons que les hypothèses supplémentaires.

Nous rappelons, tout d'abord un résultat concernant la convergence des estimateurs $\widehat{\varphi_n}(x)$ et $\widehat{\Phi_n}(x)$

Théorème 3.1 *(Convergence en moyenne quadratique de* $\widehat{\varphi_n}(x)$*) :*

Supposons $E[Y^2] < \infty$ *et posons* $v(\cdot) = Var[Y \mid X = \cdot]$,
Si

- $\varphi(\cdot)$ *est continue au point* x

Alors $\widehat{\varphi_n}(x)$ *converge en moyenne quadratique vers* $\varphi(x)$.

Si de plus :

- $f(\cdot)$ *et* $v(\cdot)$ *sont continues au point* x

Alors $\widehat{\Phi_n}(x)$ *converge en moyenne quadratique vers* $\Phi(x)$.

La démonstration de ce théorème découle de la définition de l'erreur quadratique de $\widehat{\varphi_n}(x)$ et du lemme de Bochner (voir Bosq et Lecoutre [12]). Cette erreur a été étudiée notamment par Collomb [17], voir également Lavergne [58] pour l'étude des moments conditionnels $E[Y^a \mid X]$ lorsque $a \in \mathbf{N}$.

Ce théorème permet de vérifier que l'estimateur $\widehat{f_n}(x)$ est un estimateur convergent de l'espérance conditionnelle $E[Y \mid X = x]$, comme l'indique le corollaire suivant.

Corollaire 3.2 *(convergence simple en probabilité) :*

Sous les hypothèses du théorème précédent et :

Si $\varphi(x) \neq 0$, *alors :*

$$\widehat{f_n}(x) \xrightarrow{p} f(x).$$

3.4.2 Propriétés de convergence uniforme de $\widehat{f_n}$

La formulation du théorème de convergence uniforme que nous reproduisons ici est tirée de l'ouvrage de Györfi, Härdle, Sarda et Vieu ([44] pp. 24-30), choisie pour la simplicité des hypothèses. Cette formulation nous donne explicitement la vitesse de convergence, qui nous sera utile dans le chapitre 4.

Théorème 3.3 *(Convergence uniforme de* $\widehat{f_n}$*) :*

Soit G *un compact de* $\Re^p$ *et* $\widehat{G}$ *un voisinage de ce compact* ($G \subset \widehat{G}$),

Supposons $E[Y^2] < \infty$ *et posons* $\sigma^2(\cdot) = Var[Y \mid X = \cdot]$, *sous les hypothèses suivantes :*

- *La densité* $\varphi(x) > 0$, $\forall x \in G$
- $\forall x \in \widehat{G}, \quad \sigma^2(x) < \infty$
- $\varphi(\cdot)$ *et* $f(\cdot)$ *sont d-fois continûment différentiables, et ont des dérivées bornées,*
- *le noyau* K *est de Parzen-Rosenblatt d'ordre* d

et si la fenêtre h *est telle que* V_n

$$V_n = h^d + \sqrt{\frac{\log(n)}{n \cdot h^p}}$$

vérifie $V_n \xrightarrow{n \to \infty} 0$, *alors*

$$\sup_{x \in G} \left| \widehat{f_n}(x) - f(x) \right| = O_p(V_n)$$

La preuve de ce résultat est donné par Györfi et alii dans le cadre de processus φ-mélangeant et n'est pas reproduite ici, nous en proposons toutefois un squelette, qui nous permettra d'obtenir un résultat sur la convergence uniforme de $\widehat{\varphi_n}(x)$.

Squelette de la démonstration :

L'estimateur de la régression s'écrit :

$$\widehat{f_n}(x) = \frac{\widehat{\Phi_n}(x)}{\widehat{\varphi_n}(x)}$$

nous pouvons décomposer $\widehat{f_n}(x) - f(x)$ sous la forme d'une somme de quatre termes :

$$\widehat{f_n}(x) - f(x) = (A + B + f(x)C + f(x)D) \cdot (\widehat{\varphi_n}(x))^{-1}$$

où :

$$\begin{aligned} A &= \widehat{\Phi_n}(x) - E\left[\widehat{\Phi_n}(x)\right] \\ B &= E\left[\widehat{\Phi_n}(x)\right] - \Phi(x) \\ C &= \varphi(x) - E\left[\widehat{\varphi_n}(x)\right] \\ D &= E\left[\widehat{\varphi_n}(x)\right] - \widehat{\varphi_n}(x) \end{aligned}$$

Sous les hypothèses enoncées,la fonction f est bornée sur G et l'estimateur $\widehat{\varphi_n}$ est presque-sûrement positif . On montre ensuite que :

$$\sup_{x \in G} A = O_p\left(\sqrt{\frac{\log(n)}{n \cdot h^p}}\right) \tag{3.3}$$

$$\sup_{x \in G} B = O_p\left(h^d\right) \tag{3.4}$$

$$\sup_{x \in G} C = O_p\left(h^d\right) \tag{3.5}$$

$$\sup_{x \in G} D = O_p\left(\sqrt{\frac{\log(n)}{n \cdot h^p}}\right) \tag{3.6}$$

Les démonstrations de 3.4 et 3.5 figurent dans Härdle et Luckhaus [48], le terme D peut être vu comme un cas particulier de A, dans lequel les Y_i sont tous égaux à 1. La démonstration de 3.6 sera donc immédiate une fois 3.3 démontré.

Pour cela l'estimateur $\widehat{\Phi_n}(x)$ est décomposé en $\Phi_n^+(x)$ et $\Phi_n^-(x)$:

$$\Phi_n^+(x) = \frac{1}{n \cdot h^p} \sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_n}\right) \mathbf{1}_{\{|Y_i| \geq M_n\}}$$

et

$$\Phi_n^-(x) = \widehat{\Phi}_n(x) - \Phi_n^+(x)$$

où $M_n = n^\xi$ est une suite croissante.
Le résultat provient du fait que pour tout $\varepsilon_0 > 0$

$$\sum_{n=1}^{\infty} \Pr\left(M_n \cdot n \cdot h^p \sup_{x \in \Re^p} \left| \Phi_n^+(x) - E\left[\Phi_n^+(x)\right] \right| > \varepsilon_0 \right) < \infty$$

et d'un lemme démontrant que $\forall \varepsilon > 0$:

$$\sum_{n=1}^{\infty} \Pr\left(V_n \cdot \sup_{x \in G} \left| \Phi_n^-(x) - E\left[\Phi_n^-(x)\right] \right| > \varepsilon \right) < \infty$$

Ce qui permet de conclure par addition des termes. □

Nous pouvons remarquer que l'addition des termes C et D nous donne :

Corollaire 3.4 *Sous les hypothèses du théorème précédent :*

$$\sup_{x \in G} |\widehat{\varphi_n}(x) - \varphi(x)| = O_p\left(V_n\right)$$

Nous utiliserons ces résultats pour l'étude non-paramétrique de nos statistiques d'enveloppement. Bosq et Lecoutre [12] nous donnent d'autres résultats de convergence de l'estimateur $\widehat{f_n}$ suivant le type de norme considérée pour mesurer l'écart de $\widehat{f_n}$ à f. Des résultats plus complets sur la convergence uniforme sont donnés également par Sarda et Vieu [76] (voir également Bierens [6]).

3.4.3 Distribution limite

Nous donnons ici le résultat principal concernant la distribution asymptotique de $\widehat{f_n}$. Ce résultat a été obtenu par Schuster [78] pour le cas univarié et à Collomb [18] dans le cas mutidimensionnel.

Des hypothèses supplémentaires sont nécessaires à ce résultat et sont similaires à celles rencontrées usuellement. De plus celles-ci sont explicitement utilisées dans la démonstration, ce qui rend leur interprétation plus facile.

Théorème 3.5 *(Normalité asymptotique de $\widehat{f_n}$) :*

Sous les hypothèses suivantes :

- $\exists\ \delta$ *tel que la fonction* $\sigma^{2+\delta}(x) \cdot \varphi(x)$ *est uniformément bornée*
- *Les fonctions* $f(x)^2 \cdot \varphi(x)$ *et* $\sigma^2(x) \cdot \varphi(x)$ *sont continues et uniformément bornées*

- *Les fonctions $\varphi(x)$ et $f(x)\cdot\varphi(x)$, ainsi que leurs dérivées et dérivées secondes sont continues et uniformément bornées.*

- *Le noyau K est de Parzen Rosenblatt d'ordre 2*

- $\varphi(x) > 0$

On a,

$$\textit{Si} \qquad h_n^2\sqrt{n\cdot h_n^p} \overset{n\to\infty}{\longrightarrow} l \qquad \textit{avec} \quad 0 \leq l < \infty$$

Alors,

$$\sqrt{n\cdot h_n^p}\left(\widehat{f_n}(x) - f(x)\right) \xrightarrow{d} \mathcal{IN}\left(\frac{l\cdot b(x)}{\varphi(x)}\ ,\ \frac{\sigma^2(x)}{\varphi(x)}\int_{\Re^p} K^2(z)\,dz\right) \tag{3.7}$$

De plus,

$$\textit{Si} \qquad h_n^2\sqrt{n\cdot h_n^p} \overset{n\to\infty}{\longrightarrow} \infty$$

Alors,

$$p\lim_{n\to\infty}\ h_n^{-2}\left(\widehat{f_n}(x) - f(x)\right) = \frac{b(x)}{\varphi(x)} \tag{3.8}$$

avec:
$b(x) = \frac{1}{2}Tr\left(\Xi\frac{\partial}{\partial x'}\frac{\partial}{\partial x'}\left[f(x)\varphi(x)\right]\right) - \frac{1}{2}f(x)Tr\left(\Xi\frac{\partial}{\partial x'}\frac{\partial}{\partial x'}\left[\varphi(x)\right]\right)$

– *où :* $\Xi = \int x'xK(x)dx$.

Ce dernier résultat pouvant également être interprété comme la convergence en distribution vers une loi dégénérée.

La démonstration de ce résultat par Bierens [7] est fort instructive et permet une décomposition intéressante entre termes "*asymptotiquement normaux*" et termes "*générant du biais*". Nous proposons ici un squelette de cette démonstration dont l'intégralité est rapportée en annexe.

Squelette de la démonstration :

La multiplication par $\widehat{\varphi}(x)$, supposé non nul, nous permet d'obtenir une expression plus simple :

$$\begin{aligned}\left(\widehat{f_n}(x) - f(x)\right)\cdot\widehat{\varphi}(x) &= \tfrac{1}{nh_n^p}\textstyle\sum_i Y_i K\left(\tfrac{X_i - x}{h_n}\right) - f(x)\left(\tfrac{1}{nh_n^p}\textstyle\sum_i K\left(\tfrac{X_i - x}{h_n}\right)\right)\\ &= \tfrac{1}{nh_n^p}\textstyle\sum_i \left(Y_i - f(x)\right)K\left(\tfrac{X_i - x}{h_n}\right)\end{aligned}$$

Nous pouvons alors décomposer cette quantité en trois termes dont les comportements asymptotiques seront différents :

$$\left(\widehat{f_n}(x) - f(x)\right) \cdot \widehat{\varphi}(x) = \frac{1}{nh_n^p} \sum_i \left(Y_i - f(x_i)\right) K\left(\frac{X_i - x}{h_n}\right)$$

$$+ \frac{1}{nh_n^p} \sum_i \left(f(x_i) - f(x)\right) K\left(\frac{X_i - x}{h_n}\right) - E\left[\frac{1}{nh_n^p} \sum_i \left(f(x_i) - f(x)\right) K\left(\frac{X_i - x}{h_n}\right)\right]$$

$$+ E\left[\frac{1}{nh_n^p} \sum_i \left(f(x_i) - f(x)\right) K\left(\frac{X_i - x}{h_n}\right)\right]$$

$$= \widehat{q_1}(x) + \widehat{q_2}(x) + \widehat{q_3}(x)$$

Le premier terme nous donne la normalité asymptotique :

- $\sqrt{n \cdot h_n^p} \cdot \widehat{q_1}(x) \xrightarrow{d} \mathcal{N}\left(0 \ , \ \frac{\sigma^2(x)}{\varphi(x)} \int K^2(t)\, dt\right)$
  Le second disparaît asymptotiquement puisque :
- $\lim_{n\to\infty} E\left[\sqrt{n \cdot h_n^p} \cdot \widehat{q_2}(x)\right]^2 = 0$
  Le dernier nous donne le biais, qui s'exprime en fonction de l.
- $\sqrt{n \cdot h_n^p} \cdot \widehat{q_3}(x) \overset{n\to\infty}{\longrightarrow} l \cdot b(x)$

L'addition de ces trois termes, nous donne le résultat puisque sous ces hypothèses, on a également $\widehat{\varphi}(x) \to \varphi(x)$. □

Ce résultat nous donne, de manière explicite, la vitesse de convergence h_n permettant de "*tuer le biais*". Cette vitesse est telle que $l = 0$, c'est-à-dire que h_n doit vérifier :

$$h_n^2 \cdot \sqrt{n \cdot h_n^p} \overset{n\to\infty}{\longrightarrow} 0$$

Soit encore :

Hypothèse 3.4 *: La fenêtre h_n vérifie $n \cdot h_n^{p+4} \overset{n\to\infty}{\longrightarrow} 0$*

Cette hypothèse peut être affinée en fonction de la régularité supposée des fonctions $f(x), \varphi(x)$ et $f(x) \cdot \varphi(x)$. Si ces fonctions sont d-fois continûment différentiables et de dérivées bornées, alors, en utilisant un noyau de classe d, le théorème sera vérifié.

L'hypothèse suivante :

Hypothèse 3.4 (bis) : La fenêtre h_n vérifie $n \cdot h_n^{p+2d} \overset{n\to\infty}{\longrightarrow} 0$

remplacera l'hypothèse 3.4 pour tuer le biais généré par $\widehat{q_3}(x)$.

Remarque :

Bierens [7] propose également une autre technique pour "*tuer le biais*" généré par $\widehat{q_3}(x)$. Il s'agit de créer un estimateur $\widetilde{f_n}$ basé sur la différence de

deux estimateurs, l'un générant du biais lié à la fenêtre h_1, l'autre estimant ce biais, par le choix approprié de la fenêtre h_2. La différence des deux estimateurs donne alors l'estimateur :

$$\widetilde{f_n}(x) = \frac{\widehat{f_1}(x) - l_1 \cdot \widehat{f_2}(x)}{1 - l_1}$$

où $\widehat{f_1}(x)$ est un estimateur classique utilisant la fenêtre h_1 générant un biais $l_1 \cdot \frac{b(x)}{\varphi(x)}$ conformément à (3.7)

et $\widehat{f_2}(x)$ est un estimateur de ce biais $\frac{b(x)}{\varphi(x)}$ utilisant la fenêtre h_2 conforme à (3.8).

Le biais est ainsi "*tué dans l'oeuf" ;* l'utilisation de cet estimateur permet donc de s'affranchir de l'hypothèse 3.4(bis). Un estimateur similaire est proposé par Härdle, Hall et Marron [47], sous le nom de "*double smoothing*".

Pour d'autres propriétés de convergence nous renvoyons à Bosq et Lecoutre [12], Collomb [18], Härdle [46] ainsi qu'à Robinson [74] ou Youndje [88], pour les propriétés de convergence de l'estimateur de la densité conditionnelle.

3.5 Fenêtres

"However there is a price to be paid for the great flexibility of nonparametric methods, which is that the smoothing parameter must be chosen"

J. S. Marron (1988)

Nous avons relevé dans la section précédente l'importance de la fenêtre intervenant dans l'estimation non-paramétrique, et nous nous proposons d'apporter des éléments de réponse, fournis par la littérature, à la question :

"Comment choisir la fenêtre ?"

En l'absence de technique de sélection, l'utilisateur de la statistique non-paramétrique sélectionne généralement la fenêtre au vu de la courbe, par essais successifs, cette technique visuelle ne pouvant d'ailleurs s'appliquer que pour des dimensions de régresseurs faibles ($p < 3$). L'aspect *arbitraire* de ce choix étant peu souhaitable, (en particulier dans un cadre général de comparaison de modèles), il était nécessaire de déterminer des critères de sélection *objectifs* de ce paramètre.

Une manière de s'assurer "*objectivement*" du comportement de l'estimateur en fonction de la fenêtre, est d'examiner un critère d'erreur entre la fonction de régression et son estimée. Certains de ces critères seront globaux ou locaux, ils peuvent être basés sur la prévision ou sur l'écart au sens d'une norme fonctionnelle, qui variera suivant l'utilisation ou les propriétés souhaitées de l'estimateur.

L'un des critères le plus utilisé repose sur l'Erreur Quadratique Intégrée (EQI) définie par[3] :

$$EQI(h) = \int_{\Re^p} \left(\widehat{f_n}(x) - f(x)\right)^2 \varphi(x)\varpi(x)\lambda(dx)$$

où $\varpi(\cdot)$ est une fonction de poids[4].

On peut également trouver une version empirique de l'EQI , l'Erreur Quadratique Empirique (EQE), en remplaçant la mesure de Lebesgue par la loi empirique des X_i :

$$EQE(h) = \sum_{i=1}^{n} \left(\widehat{f_n}(X_i) - f(X_i)\right)^2 \varpi(X_i)$$

ou une version globale, l'Erreur Quadratique Intégrée Moyenne ($EQIM$) obtenue en prenant l'espérance de l'EQI :

$$EQIM(h) = \mathbf{E}\left[EQI(h)\right] = \mathbf{E}\left[\int_{\Re^p} \left(\widehat{f_n}(x) - f(x)\right)^2 \varphi(x)\varpi(x)\lambda(dx)\right]$$

Härdle et Marron [50] montrent que ces trois mesures quadratiques sont asymptotiquement équivalentes pour une grande variété d'estimateurs.

La fenêtre idéale, h_{opt} doit alors réaliser le minimum de l'un de ces critères qui dépendent des fonctions inconnues f et φ. On la définit par :

$$h_{opt} = Arg \min_{h \in H_n} EQI(h)$$

Le rôle de la fenêtre h dans ces critères d'erreur peut être vu à travers une formulation de Vieu [87] donnant une évaluation asymptotique de l'EQI pour des fonctions f et φ d-fois continûment différentiables et pour des noyaux d'ordre d :

$$EQI(h) = B \cdot h^{2d} + \frac{V}{n \cdot h^p} + o_p\left(h^{2d} + \frac{1}{n \cdot h^p}\right) \tag{3.9}$$

[3] Il existe bien d'autres critères et nous ne citons ici que les plus "*populaires*", voir Härdle [46] , Härdle et Marron [50] ou Marron [64].

[4] Cette fonction de poids est généralement introduite dans ces définitions pour compenser les problèmes d'estimation lorsque la densité des régresseurs devient faible ("*Effets de bord*"). Les conditions sur cette fonction sont donc liées à la densité inconnue $\varphi(x)$, nécessitant une information supplémentaire sur cette densité.

où : B et V sont deux nombres réels finis.

Nous retrouvons ici l'aspect d'arbitrage entre *B*iais et *V*ariance joué par la fenêtre, puisque les termes B et V correspondent respectivement à des termes de biais et de variance approchés (voir également Hall [45], Rice [73], ou Härdle et Marron[50]). A propos de cette expression Härdle [46] écrit d'ailleurs :

"(· · ·) one gets a feeling of what the smoothing problem is about : Balance the variance versus the biais"

Cette expression permet de dégager deux optiques menant à deux procédures de choix de la fenêtre.

- Soit on estime la fenêtre minimisant l'$EQI(h)$ en estimant B et V,
- Soit on estime l'$EQI(h)$ et on sélectionne la fenêtre minimisant ce nouveau critère.

La première solution mène aux techniques de "*Plug-in*", la seconde aux méthodes de "*validation croisée*".

3.5.1 Le "*plug-in*"

La méthode du "Plug-in" repose sur la sélection de la fenêtre h_d minimisant l'$EQI(h)$ donné par (3.9). Cette fenêtre peut être approchée par $\widehat{h_d}$ obtenue en estimant les termes B et V qui dépendent des fonctions inconnues f , φ (et de leurs dérivées) ainsi que du noyau K.

Cette technique est très satisfaisante théoriquement puisque l'expression de h_d minimisant l'$EQI(h)$ dans l'équation (3.9) est :

$$h_d = \left(\frac{p \cdot V}{2d \cdot B}\right)^{\frac{1}{2d+p}} \cdot n^{\frac{-1}{2d+p}}$$

et que $EQI(h_d)$ est alors de l'ordre

$$EQI(h_d) = O_p\left(n^{-\frac{2d}{2d+p}}\right)$$

Cette vitesse de convergence est donnée par Stone [83] comme étant la vitesse de convergence optimale dans la classe des fonctions de régression d-fois continûment différentiables. De plus la vitesse de convergence de h_d et donc de $\widehat{h_d}$ est explicitement $n^{\frac{-1}{2d+p}}$, toutefois ces fenêtres ne vérifient pas l'hypothèse 3.4 (bis), puisque $n \cdot h_d^{2d+p} = \left(\frac{p \cdot V}{2d \cdot B}\right) \not\longrightarrow 0$.

De plus cette technique nous confine à l'étude de fonctions de régression suffisamment régulières (d-fois continûment différentiables). Enfin, des difficultés importantes se posent en pratique : pour calculer $\widehat{h_d}$ il faut, en effet, estimer les constantes B et V et donc les dérivées des fonctions f et φ ce qui s'avère techniquement délicat, hors d'un contexte de régresseurs uniformément répartis (voir Vieu [87]).

3.5.2 La "*Validation Croisée*"

L'idée de base consiste à trouver une fonction de score $CV(h)$ ayant la même structure que l'$EQI(h)$ et dont le calcul soit plus simple. On sélectionne alors la fenêtre $\widehat{h}_{cv}$ minimisant ce critère dont on attend le même comportement asymptotique que h_{opt}.

Le critère $CV(h)$ est obtenu à partir de l'Erreur Quadratique Empirique ($EQE(h)$) dans laquelle l'estimateur $\widehat{f_n}(X_i)$ est remplacé par l'estimateur de "*leave-one-out*" $\widehat{f_n^{-i}}(X_i)$ et $f(X_i)$ est estimé naïvement par Y_i (voir Härdle[46] pp.152-153).

On choisit alors $\widehat{h}_{cv} = Arg\min_{h\in H_n} CV(h)$ où :

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{f_n^{-i}}(X_i)\right)^2 \varpi(X_i)$$

et

$$\widehat{f_n^{-i}}(X_i) = \frac{\frac{1}{nh_n^p}\sum_{j\neq i} Y_i K\left(\frac{X_j - X_i}{h_n}\right)}{\frac{1}{nh_n^p}\sum_{j\neq i} K\left(\frac{X_j - X_i}{h_n}\right)}$$

Härdle et Marron [50] démontrent que la fenêtre ainsi obtenue vérifie la propriété d'optimalité asymptotique suivante :

$$\frac{EQI(\widehat{h}_{cv})}{EQI(h_{opt})} \longrightarrow 1 \quad p.s.$$

sous les hypothèses

Hypothèse 3.5 *(Optimalité asymptotique) :*

- $H_n = \left[\underline{h}, \overline{h}\right] = \left[a_n \cdot n^{-\frac{1}{p}}, a_n^{-1}\right]$

  où $a_n = C \cdot n^{\delta}$ *et* C, δ *sont des constantes positives,* $\delta \in \left[0, \frac{1}{p}\right]$

- *Les fonctions* f, φ *et* K *sont Hölder continues*[5]

[5]Une fonction g est Hölder continue s'il existe des constantes positives M et ξ, telles que :

$$|g(x) - g(t)| \leq M \cdot \|x - t\|^{\xi}$$

- *Les moments conditionnels de* $Y \mid X$ *sont bornés, c'est-à-dire :*

  $\forall\, q > 0, \quad \exists\ A_q > 0 \quad$ *tel que* $E\left[|Y|^q \mid X = x\right] < A_q\,,\ \forall\, x$

- *La fonction de poids* $\varpi(\cdot)$ *est à support compact* S

- *La densité marginale* $\varphi(x)$ *est bornée inférieurement sur l'intérieur de ce support* S

L'inconvénient principal est que la fenêtre $\widehat{h}_{cv}$, qui est ici un estimateur, présente une grande variabilité, c'est-à-dire que pour deux échantillons distincts issus de la même distribution, les fenêtres obtenues seront très différentes. Ce problème a été étudié par Härdle et Marron [50] où il est montré que $\widehat{h}_{cv}$ converge "*très lentement*" vers h_{opt}. Une technique pour pallier à cet inconvénient consiste peut-être à utiliser le "double smoothing" proposé par Härdle, Hall et Marron [47] et évoqué plus haut.

Cette méthode présente cependant de nombreux avantages : outre le fait qu'elle ne demande pour être applicable, que des hypothèses faibles sur le degré de différentiabilité de f, c'est une méthode automatique entièrement guidée par les données. Ce point est particulièrement satisfaisant dans le contexte de comparaison de modèles. Nous utiliserons d'ailleurs cette méthode dans le chapitre suivant pour éviter tout choix arbitraire pouvant influer sur la qualité des estimateurs.

3.5.3 Autres méthodes

Il existe de nombreux raffinements de la validation croisée Vieu [86] propose de sélectionner la fenêtre localement en utilisant un critère de validation croisé local, tenant compte de la densité autour de chaque observation. Ce critère est malheureusement encore assez coûteux en temps de calcul pour être utilisé en pratique

Une autre variante consiste à détruire plusieurs points en utilisant un estimateur de "*leave-several-out*" dans la définition du critère CV.

L'introduction d'une fonction pénalisante Ξ dans l'estimateur naïf de l'EQE permet également d'obtenir un critère de Score sur la base duquel est estimée la fenêtre. Härdle [46] (pp.155-167) nous donne une étude comparative sur un échantillon, de différentes fonctions pénalisantes.

D'autres méthodes sont exposées dans la revue sur ce sujet réalisée par Vieu [87], parmi lesquelles les méthodes de Bootstrap semblent également particulièrement prometteuses (voir également Härdle [46]).

3.6 Conclusion

Nous avons exposé dans ce chapitre quelques unes des méthodes d'estimation fonctionnelle de la régression. Ces méthodes permettent l'étude des modèles de régression en l'absence de forme fonctionnelle prédéfinie, et en l'absence de spécification de la loi des résidus. Cette liberté dans la spécification (ou plutôt dans l'absence de spécification) des modèles de régression n'est cependant pas exempte de règles. La sélection du paramètre de lissage, dans chacune de ces méthodes est soumis à des contraintes et les règles de sélection pratiques de ce paramètre sont encore à l'étude.

Une autre contrainte nous est donnée par l'inexistence d'un estimateur sans biais de la régression montré par Collomb [17], il en résulte une approche asymptotique nécessitant un nombre important d'observations. Ce point est aggravé par la perte d'une vitesse de convergence "*paramétrique*" (en $\sqrt{n}$), montré par Stone [83] ; la convergence non-paramétrique étant plus lente, ces méthodes exigent un plus grand nombre de données.

Nous avons choisi de développer plus particulièrement la méthode du noyau pour de simples raisons : cette méthode est la plus développée à ce jour, et les propriétés des estimateurs sont maintenant bien connues. En outre elle bénéficie d'une abondante littérature sur des problèmes théoriques et appliqués. Enfin, une procédure d'estimation de la fenêtre est possible dans le cadre de cette méthode. Cette procédure est entièrement guidée par les données et présente donc un caractère "*objectif*" particulièrement appréciable dans le cadre de comparaison de modèles. Nous utiliserons cette méthode (et cette procédure) pour l'estimation des fonctions de régression dans le chapitre suivant.

3.7 Annexe au chapitre 3

Notations

Les "petits-o" et les "grands-O", que l'on trouve couramment dans la littérature sont rappelés de manière précise ici. Ces symboles ont été introduits par Landau pour simplifier les relations entre quantités (stochastiques ou non) de même Ordre de grandeur, ou d'un $o$rdre de grandeur inférieur asymptotiquement. Nous nous servirons de ces notations dans les démonstrations à venir.

Définition 3.4 *: Si f et g sont deux fonctions réelles de la variable entière n, alors la notation $f(n) = o\left(g(n)\right)$ signifie que :*

$$\lim_{n\to\infty} \left(\frac{f(n)}{g(n)}\right) = 0$$

Il est important de noter que $g(n)$ peut avoir n'importe quel comportement lorsque $n \to \infty$, en particulier la notation $f(n) = o\left(1\right)$ signifie simplement que la suite $f(n) \to 0$ lorsque $n \to \infty$.

Définition 3.5 *: Si f et g sont deux fonctions réelles de la variable entière n, alors la notation $f(n) = O\left(g(n)\right)$ signifie qu'il existe une constante $K > 0$, indépendante de n, et un entier N tels que :*

$$\left|\frac{f(n)}{g(n)}\right| < K \quad , \quad \forall n > N$$

Ceci signifie donc que f et g ont le "*même ordre de grandeur*"[6]

De même, des relations liant les ordres de grandeur de quantités stochastiques sont exprimées par les célèbres "petits-o_p"et "grands-O_p" définis comme suit.

Définition 3.6 *: Si a_n est une suite de variables aléatoires et g est une fonction réelle de la variable entière n, alors la notation $a_n = o_p\left(g(n)\right)$ signifie que :*

$$p\lim_{n\to\infty} \left(\frac{a_n}{g(n)}\right) = 0$$

De manière similaire, la notation $a_n = O_p\left(g(n)\right)$ signifie que il existe une constante $K > 0$, telle que $\forall \varepsilon > 0$, $\exists$ un entier N_ε tel que :

$$P_r\left(\left|\frac{a_n}{g(n)}\right| > K\right) < \varepsilon \quad , \quad \forall n > N_\varepsilon$$

[6]Cette définition n'exclut pas la possible nullité de ce rapport, l'expression "*de même ordre que*" peut être trompeuse.

Démonstration du théorème 3.5 :

Classiquement, nous écrivons la différence $\left(\widehat{f_n}(x) - f(x)\right)$sous la forme :

$$\begin{aligned}\left(\widehat{f_n}(x) - f(x)\right) &= \frac{\frac{1}{nh_n^p}\sum_i Y_i K\left(\frac{X_i - x}{h_n}\right)}{\frac{1}{nh_n^p}\sum_i K\left(\frac{X_i - x}{h_n}\right)} - f(x)\frac{\left(\frac{1}{nh_n^p}\sum_i K\left(\frac{X_i - x}{h_n}\right)\right)}{\left(\frac{1}{nh_n^p}\sum_i K\left(\frac{X_i - x}{h_n}\right)\right)} \\ &= \frac{\frac{1}{nh_n^p}\sum_i (Y_i - f(x)) K\left(\frac{X_i - x}{h_n}\right)}{\left(\frac{1}{nh_n^p}\sum_i K\left(\frac{X_i - x}{h_n}\right)\right)}\end{aligned}$$

Nous pouvons simplifier l'écriture en multipliant les deux membres par l'estimateur $\widehat{\varphi}(x)$, nous obtenons ainsi la décomposition de $\left(\widehat{f_n}(x) - f(x)\right) \cdot \widehat{\varphi}(x)$:

$$= \frac{1}{nh_n^p}\sum_i (Y_i - f(x_i)) K\left(\frac{X_i - x}{h_n}\right)$$

$$+ \frac{1}{nh_n^p}\sum_i (f(x_i) - f(x)) K\left(\frac{X_i - x}{h_n}\right) - E\left[\frac{1}{nh_n^p}\sum_i (f(x_i) - f(x)) K\left(\frac{X_i - x}{h_n}\right)\right]$$

$$+ E\left[\frac{1}{nh_n^p}\sum_i (f(x_i) - f(x)) K\left(\frac{X_i - x}{h_n}\right)\right]$$

$$= \widehat{q_1}(x) + \widehat{q_2}(x) + \widehat{q_3}(x)$$

Comme mentionné plus haut, les trois termes ont des comportements asymptotiques différents, que nous analyserons séparément en trois parties :

Premier terme : $\sqrt{nh_n^p} \cdot \widehat{q_1}(x) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(x)\varphi(x) \int K^2(z)\, dz\right)$

Nous pouvons écrire $\sqrt{n \cdot h_n^p} \cdot \widehat{q_1}(x)$ sous une forme permettant d'appliquer le théorème central limite de Lyapunov, voir Serfling [79] :

$$\sqrt{n \cdot h_n^p} \cdot \widehat{q_1}(x) = \frac{1}{\sqrt{n}} \sum_i v_{n,i}(x)$$

avec $v_{n,i}(x) = \frac{1}{\sqrt{h_n^p}} \cdot u_i K\left(\frac{X_i - x}{h_n}\right)$, où $u_i = (Y_i - f(x_i)$

On a alors

- $E\left[v_{n,i}(x)\right] = 0$
- Comportement de $E\left[v_{n,i}(x)^2\right]$

$$\begin{aligned} E\left[v_{n,i}(x)^2\right] &= \tfrac{1}{h_n^p}\int u_i^2 K^2\left(\tfrac{x_i-x}{h_n}\right)\varphi(x_i,y_i)\,dx_i dy_i \\ &= \tfrac{1}{h_n^p}\int u_i^2 K^2\left(\tfrac{x_i-x}{h_n}\right)\varphi(y_i \mid x_i)\varphi(x_i)\,dx_i dy_i \\ &= \tfrac{1}{h_n^p}\int \sigma^2(x_i) K^2\left(\tfrac{x_i-x}{h_n}\right)\varphi(x_i)\,dx_i \end{aligned}$$

Le changement de variable $z = \frac{x-x_i}{h_n}$ nous donne,

$$E\left[v_{n,i}(x)^2\right] = \int \sigma^2(x - zh_n) K^2(z)\,\varphi(x - zh_n)\,dz$$

La fonction $\sigma^2(x)\varphi(x)$ étant continue et uniformément bornée, le théorème de la convergence bornée (voir par exemple Metivier [63]) s'applique :

$$E\left[v_{n,i}(x)^2\right] = \int \sigma^2(x - zh_n) K^2(z)\,\varphi(x - zh_n)\,dz \overset{h_n\to 0}{\longrightarrow} \sigma^2(x)\varphi(x)\int K^2(z)\,dz$$

Ce terme détermine la variance asymptotique de $\sqrt{n\cdot h_n^p}\cdot\widehat{q_1}(x)$.

- Nous devons nous assurer du comportement de :

$$\sum_i E\left[\left(\frac{|v_{n,i}(x)|}{\sqrt{n}}\right)^{2+\delta}\right] = \left(\frac{1}{\sqrt{nh_n^p}}\right)^{\delta} E\left[|u_i|^{2+\delta}\left|K\left(\frac{x_i-x}{h_n}\right)\right|^{2+\delta} h_n^{-p}\right]$$

par la même technique :

$$\begin{aligned} \textstyle\sum_i E\left[\left(\frac{|v_{n,i}(x)|}{\sqrt{n}}\right)^{2+\delta}\right] &= \left(\tfrac{1}{\sqrt{nh_n^p}}\right)^{\delta}\int \sigma^{2+\delta}(x - zh_{ni})\varphi(x - zh_n)\,|K(z)|^{2+\delta}\,dz \\ &= O_p\left(\tfrac{1}{\sqrt{nh_n^p}}\right)^{\delta} \overset{n\to\infty}{\longrightarrow} 0 \quad pour \quad \delta > 0 \end{aligned}$$

Le théorème central limite de Lyapounov s'applique donc et,

$$\frac{1}{\sqrt{n}}\sum_i v_{n,i}(x) \overset{d}{\longrightarrow} \mathcal{N}\left(0, \sigma^2(x)\varphi(x)\int K^2(z)\,dz\right)$$

ce qui termine l'étude du premier terme .

Deuxième terme : $E\left[\left(\sqrt{nh_n^p}\cdot\widehat{q_2}(x)\right)^2\right] \overset{n\to\infty}{\longrightarrow} 0$

$$\widehat{q_2}(x) = \frac{1}{nh_n^p}\sum_{i=1}^{n}\left\{(f(x_i) - f(x))\,K\left(\frac{x_i-x}{h_n}\right) - E\left[(f(x_i) - f(x))\,K\left(\frac{x_i-x}{h_n}\right)\right]\right\}$$

De même que précédemment,

$$E\left[\left(\sqrt{nh_n^p}\cdot\widehat{q_2}(x)\right)^2\right] = \int\left(f(x-zh_n)-f(x)\right)^2\varphi(x-zh_n)K^2(z)\,dz$$

$$-\tfrac{1}{h_n^p}\left\{\int\left(f(x-zh_n)-f(x)\right)\varphi(x-zh_n)\,K(z)\,dz\right\}^2$$

$$\stackrel{n\to\infty}{\longrightarrow} 0$$

par convergence bornée.

Troisième terme : $\sqrt{nh_n^p}\cdot\widehat{q_3}(x)\longrightarrow l\cdot b(x)$

Nous utiliserons ici la formule de Taylor pour une fonction G(x) deux fois continûment différentiable, à savoir :

$\exists\lambda_n\in[0,1]$ tel que

$$G(x)-G(x-zh_n) = -h_n z'\tfrac{\partial}{\partial x}\left[G(x)\right]$$

$$+\tfrac{1}{2}\;h_n^2\;z'\left[\tfrac{\partial^2}{\partial x\partial x'}G(x-\lambda_n h_n\,z)\right]\;z$$

On a :

$$\widehat{q_3}(x) = \tfrac{1}{nh_n^p}\textstyle\sum_{i=1}^n E\left[\left(f(x_i)-f(x)\right)K\left(\tfrac{x_i-x}{h_n}\right)\right]$$

$$= \tfrac{1}{h_n^p}\int\left(f(x_i)-f(x)\right)K\left(\tfrac{x_i-x}{h_n}\right)\varphi(x_i)dx_i$$

En opérant le même changement de variable que précédemment,

$$\widehat{q_3}(x) = \int\left(f(x-zh_n)-f(x)\right)K(z)\,\varphi(x-zh_n)dz$$

nous pouvons ajouter et retrancher $(f(x)\varphi(x))$ aux deux membres

$$\widehat{q_3}(x) = \int\left(f(x-zh_n)\varphi(x-zh_n)-f(x)\varphi(x)\right)K(z)\,dz$$

$$-\int\left(\varphi(x-zh_n)-\varphi(x)\right)f(x)K(z)\,dz$$

En appliquant la formule de Taylor aux fonctions deux fois différentiables $f(x)\varphi(x)$ et $\varphi(x)$:

$$\widehat{q_3}(x) = -h_n\int z'\tfrac{\partial}{\partial x}\left[f(x)\varphi(x)\right]K(z)\,dz$$

$$+\tfrac{1}{2}h_n^2\int z'\left[\tfrac{\partial^2}{\partial x\partial x'}f(x-\lambda_n\cdot zh_n)\varphi(x-\lambda_n h_n\cdot z)\right]\;z\,K(z)\,dz$$

$$+h_n f(x)\int z'\tfrac{\partial}{\partial x}\left[\varphi(x)\right]K(z)\,dz$$

$$-\tfrac{1}{2}h_n^2 f(x)\int z'\left[\tfrac{\partial^2}{\partial x\partial x'}\varphi(x-\lambda_n h_n\cdot z)\right]\;z\,K(z)\,dz$$

soit encore,

$$\begin{aligned}\widehat{q_3}(x) &= -\, h_n \tfrac{\partial}{\partial x}\left[f(x)\varphi(x)\right] \int z' K\left(z\right) dz \\ &\quad + \tfrac{1}{2} h_n^2 \int z' \left[\tfrac{\partial^2}{\partial x \partial x'} f(x - \lambda_n \cdot z h_n)\varphi(x - \lambda_n h_n \cdot z)\right] \; z\, K\left(z\right) dz \\ &\quad + h_n f(x) \tfrac{\partial}{\partial x}\left[\varphi(x)\right] \int z' K\left(z\right) dz \\ &\quad - \tfrac{1}{2} h_n^2 f(x) \int z' \left[\tfrac{\partial^2}{\partial x \partial x'} \varphi(x - \lambda_n h_n \cdot z)\right] \; z\, K\left(z\right) dz\end{aligned}$$

Le noyau K est d'ordre 2, donc $\int z' K\left(z\right) dz = 0$ et l'on pose :

$$\int z' z K(z) dz = \Xi$$

et donc,

$$\begin{aligned}\widehat{q_3}(x) &= \tfrac{1}{2}\; h_n^2 \;\int z' \left[\tfrac{\partial^2}{\partial x \partial x'} f(x - \lambda_n \cdot z h_n)\varphi(x - \lambda_n h_n \cdot z)\right] \; z\, K\left(z\right) dz \\ &\quad - \tfrac{1}{2}\; h_n^2 \;\; f(x) \int z' \left[\tfrac{\partial^2}{\partial x \partial x'} \varphi(x - \lambda_n h_n \cdot z)\right] \; z\, K\left(z\right) dz\end{aligned}$$

Les dérivées des fonctions $f(x)\varphi(x)$ et $\varphi(x)$ sont bornées ce qui permet d'appliquer une nouvelle fois le théorème de la convergence bornée :

$$\begin{aligned}h_n^{-2} \cdot \widehat{q_3}(x) \;\longrightarrow\; & \tfrac{1}{2} \int z' \; z\, K\left(z\right) dz \cdot \left[\tfrac{\partial^2}{\partial x \partial x'} f(x)\varphi(x)\right] \\ & - \tfrac{1}{2}\; f(x) \int z' \; z\, K\left(z\right) dz \cdot \left[\tfrac{\partial^2}{\partial x \partial x'} \varphi(x)\right] \\ & = \tfrac{1}{2} Tr\; \left\{\Xi \cdot \left[\tfrac{\partial^2}{\partial x \partial x'}\; f(x)\varphi(x)\right]\right\} - \tfrac{1}{2} Tr\left\{\Xi \cdot f(x) \left[\tfrac{\partial^2}{\partial x \partial x'} \varphi(x)\right]\right\} \\ & = b(x)\end{aligned}$$

c'est-à-dire que $\sqrt{n h_n^p} \cdot \widehat{q_3}(x) \longrightarrow l \cdot b(x)$.

L'addition des trois termes nous donne le résultat. □

[12pt,qqa4lrep]report

Chapter 4

Procédures paramétriques et non-paramétriques

4.1 Introduction

La notion d'enveloppement introduite par Mizon et Richard [66], et développée dans la première partie, est ici élargie à l'étude de modèles de régression "*libres*" de toute forme prédéfinie. Cette étude nous amènera à considérer des modèles de régression munis d'estimateurs non-paramétriques, définis dans le chapitre précédent, ainsi que des modèles paramétriques standards.

Nous proposerons différents tests concernant l'enveloppement d'un modèle de régression $\mathcal{M}_2$ basé sur la variable Z, par un modèle $\mathcal{M}_1$ ayant pour variable conditionnante X, nous plaçant alors principalement dans un cadre de régresseurs non-emboîtés.

Il est important de distinguer les modèles linéaires qui répondent à une modélisation particulière, des modèles libres de toute forme fonctionnelle sur lesquels on opère une approximation linéaire. En effet, l'opérateur de projection dans un espace L^2 nous donne une approximation linéaire d'un modèle de régression, indépendamment de la linéarité du modèle lui-même.

Nous utiliserons section (4.2) différents opérateurs de projection afin de retrouver les résultats de Mizon et Richard [66] dans un contexte plus général.

Le point clé de notre analyse repose sur l'indépendance des choix de régresseurs vis-à-vis du choix de la forme des modèles de régressions. Autrement formulé ce problème pose la question suivante :

"L'exclusion de la variable Z dans le modèle $\mathcal{M}_1$ est elle robuste au choix de la forme fonctionnelle des modèles de régression ? "

Nous proposerons différents tests concernant l'enveloppement d'un modèle $\mathcal{M}_2$ par un modèle $\mathcal{M}_1$ en étudiant les spécifications paramétriques et non-

		Modèle M_2	
Modèle M_1		Paramétrique	Non-paramétrique
	Paramétrique	**PP**	**PN**
	Non-paramétrique	**NP**	**NN**

Table 4.1: Les 4 cas

paramétriques pour chacun des modèles. Quatre situations se présentent et seront notées conformément à la table 4.1. Pour chacune de ces situations, nous proposerons section 4.3 une statistique de test d'enveloppement permettant de répondre à cette question.

4.2 Notations et modèles

Nous définissons tout d'abord les observations comme étant n réalisations indépendantes du vecteur aléatoire $S = (Y, X, Z)$ et notées $(S_i)_{i=1,\dots,n}$ où $Y_i \in \Re$, $X_i \in \Re^p$ et $Z_i \in \Re^q$. Essentiellement les variables X et Z représentent les variables exogènes associées aux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ respectivement.

Formellement, nous supposerons que $S_i = (Y_i, X_i, Z_i)_{i=1,\dots,n}$ constitue un processus centré, *iid*, de carré intégrable défini sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P}_0)$. Il est caractérisé par la densité inconnue $\varphi(S_i)$ par rapport à la mesure de Lebesgue sur $\Re^{p+q+1}$. La probabilité $\mathcal{P}_0$ est évidemment inconnue et nous limiterons notre attention à l'étude de paramètres ou de fonctions définies à partir de $\mathcal{P}_0$.

Les composantes de (X_i, Z_i) sont supposées linéairement indépendantes. Cette dernière hypothèse peut être relâchée, et les vecteurs X_i et Z_i pourront éventuellement être imbriqués, dans ce cas la densité φ sera considérée par rapport à la mesure de Lebesgue sur un sous-espace de $\Re^{p+q+1}$, nous ne détaillerons toutefois pas ce cas.

Nous utiliserons les notations f, g et r pour représenter les espérances conditionnelles suivantes, dont les définitions sont conformes à celles données section (2.2.1) :

$$f(x) = E[Y \mid X = x]$$

$$g(z) = E[Y \mid Z = z]$$

et

$$r(x, z) = E[Y \mid X = x, Z = z]$$

Il est important de reconnaître que les approximations linéaires peuvent être utilisées sans que les fonctions de régression ne soient elles-mêmes linéaires. Nous noterons $L(\cdot \mid \cdot)$ les projections définies comme suit :

Définition 4.1 *(Projections dans L^2)*
La projection de Y_i sur le sous espace engendré par les X_i est[1] *:*

$$L(Y_i \mid X_i) = \beta' X_i \quad avec \quad \beta = \left(E\left[X_i X_i'\right]\right)^{-1} E\left[X_i Y_i\right]$$

Le vecteur de paramètre β est alors une fonction à valeur dans $\Re^p$ de la densité inconnue $\varphi(\cdot)$, et donc de $\mathcal{P}_0$.

On définit de même la projection sur les Z_i par :

$$L(Y_i \mid Z_i) = \gamma' Z_i \quad avec \quad \gamma = \left(E\left[Z_i Z_i'\right]\right)^{-1} E\left[Z_i Y_i\right]$$

et la projection sur l'espace engendré par (X_i, Z_i),

$$L(Y_i \mid X_i, Z_i) = \alpha' W_i \quad avec \quad \alpha = \left(E\left[W_i W_i'\right]\right)^{-1} E\left[W_i Y_i\right]$$

où W_i est une base de l'espace engendré par (X_i, Z_i). Nous supposerons par la suite que $W_i' = (X_i, Z_i)$, X_i et Z_i étant strictement non-emboîtés.

4.2.1 Hypothèses

Le processus $S_i = (Y_i, X_i, Z_i)_{i=1,\ldots,n}$ étant de carré intégrable, les fonctions f, g et r sont elles-mêmes de carré intégrable sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathcal{P}_0)$.

Les conditions de régularité suivantes sont supposées pour ces fonctions :

Hypothèse 4.1 *(Version continue des fonctions de régression).*
Il existe une version continue des fonctions f, g et r ainsi que des densités, densités marginales et conditionnelles (représentées par la même fonction φ).

Cette hypothèse sera maintenue tout au long de ce travail.

Nous chercherons à valider le modèle $\mathcal{M}_1$, qui sera le modèle "*enveloppant*", en utilisant le modèle "*à envelopper*" $\mathcal{M}_2$. Le modèle $\mathcal{M}_1$ est basé sur l'exclusion de la variable Z, et pourra présenter deux aspects différents suivant l'hypothèse d'exclusion de cette variable.

Hypothèse 4.2 *:*
L'exclusion de la variable Z du modèle $\mathcal{M}_1$ peut être considérée par deux hypothèses distinctes :

$$\mathcal{H}_1 \quad : \quad E\left[Y \mid X, Z\right] = E\left[Y \mid X\right],$$

[1] Nous représenterons l'espérance relative à $\mathcal{P}_0$ par la lettre générique E.

ce qui correspond à une hypothèse d'indépendance de l'espérance conditionnelle, ou par l'hypothèse :

$$\mathcal{H}_2 \quad : \quad L[Y \mid X, Z] = L[Y \mid X] \ ,$$

qui est une condition d'orthogonalité conditionnelle (ou d'indépendance linéaire).

Comme nous l'avons déjà mentionné $\mathcal{H}_2$ ne signifie pas que la fonction de régression est linéaire. La linéarité de cette fonction correspond à une troisième hypothèse :

$$\mathcal{H}_3 \quad : \quad E[Y \mid X, Z] = L[Y \mid X, Z]$$

Nous utiliserons également, mais plus rarement, une dernière hypothèse concernant le carré de Y :

$$\mathcal{H}_4 \quad : \quad E\left[Y^2 \mid X, Z\right] = E\left[Y^2 \mid X\right]$$

Remarques :

Quelques propriétés simples découlent de la combinaison des hypothèses précédentes :

- Le couple d'hypothèses $(\mathcal{H}_1, \mathcal{H}_4)$ implique l'égalité des variances conditionnelles :
$$V(Y \mid X, Z) = V(Y \mid X)$$

- Le couple $(\mathcal{H}_2, \mathcal{H}_4)$ entraîne l'égalité des variances des "*résidus linéaires*" :

$$E\left[(Y - L(Y \mid X))^2 \mid X, Z\right] = E\left[(Y - L(Y \mid X))^2 \mid X\right]$$

De manière évidente, si S est normalement distribué alors $\mathcal{H}_3$ est vérifié et l'hypothèse $\mathcal{H}_2$ est équivalente à $\mathcal{H}_1$.

4.2.2 Modèles

Dans une optique non-paramétrique, le modèle $\mathcal{M}_1$ "libre" sera caractérisé par les hypothèses $\mathcal{H}_1$ ou $(\mathcal{H}_1, \mathcal{H}_4)$, tandis que la linéarité (paramétrique) sera caractérisée par $(\mathcal{H}_2, \mathcal{H}_4)$ ou $(\mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4)$, ces deux dernières combinaisons correspondant respectivement à la linéarité faible ou forte[2].

[2]Dans la suite de ce chapitre, nous regrouperons la linéarité faible et forte sous le terme commun de régression linéaire. Nous serons néanmoins explicites sur la validité de $\mathcal{H}_3$ dans les cas où cette hypothèse s'avère d'importance.

Le modèle rival $\mathcal{M}_2$ sera construit avec la variable Z comme unique régresseur, nous considérerons également une version linéaire et une version *"libre"* de ce modèle.

Un troisième modèle $\mathcal{M}$ est également d'intérêt, c'est le modèle emboîtant $\mathcal{M}_1$ et $\mathcal{M}_2$ construit sur les régresseurs W_i.

Dans l'optique de $\mathcal{M}_1$, ou de son propriétaire, ces modèles ne présentent qu'un intérêt limité, puisque $\mathcal{M}_2$ est vu comme un modèle mal-spécifié et $\mathcal{M}$ comme un sur-modèle. Ces deux modèles seront les instruments de la construction de tests d'enveloppement bâtis en vue de valider le modèle $\mathcal{M}_1$.

Afin de rester cohérent avec notre notion de modèle définie dans la section 1.7**,** ces modèles sont associés à des estimateurs. Les régressions linéaires seront estimées classiquement par l'estimateur des moindres carrés, les modèles "libres" seront estimés non-paramétriquement par la méthode du noyau de convolution. Nous obtiendrons donc des estimateurs paramétriques et des estimateurs fonctionnels pour chacun de ces modèles.

Considérons tout d'abord la version linéaire de $\mathcal{M}_1$.

Un estimateur naturel de β défini en (4.1) est :

$$\widehat{\beta} = \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \cdot \sum_{i=1}^{n} X_i Y_i$$

Les estimateurs correspondants pour γ dans $\mathcal{M}_2$ et pour α dans le modèle emboîtant $\mathcal{M}$ sont :

$$\widehat{\gamma} = \left(\sum_{i=1}^{n} Z_i Z_i' \right)^{-1} \cdot \sum_{i=1}^{n} Z_i Y_i$$

et,

$$\widehat{\alpha} = \left(\sum_{i=1}^{n} W_i W_i' \right)^{-1} \cdot \sum_{i=1}^{n} W_i Y_i$$

Les estimateurs non-paramétriques des fonctions de régression f, g et r sont les estimateurs du noyau de convolution conformes à la définition de la section (3.3), à savoir :

$$\widehat{f_n}(x) = \frac{\frac{1}{nh_n^p} \sum_i Y_i K\left(\frac{X_i - x}{h_n}\right)}{\frac{1}{nh_n^p} \sum_i K\left(\frac{X_i - x}{h_n}\right)}$$

de même,

$$\widehat{g_n}(z) = \frac{\frac{1}{nk_n^q} \sum_i Y_i K\left(\frac{Z_i - z}{k_n}\right)}{\frac{1}{nk_n^q} \sum_i K\left(\frac{Z_i - z}{k_n}\right)}$$

et

$$\widehat{r_n}(x,z) = \frac{\frac{1}{nh_n^p \cdot k_n^q} \sum_i Y_i K\left(\frac{X_i - x}{h_n}, \frac{Z_i - z}{k_n}\right)}{\frac{1}{nh_n^p \cdot k_n^q} \sum_i K\left(\frac{X_i - x}{h_n}, \frac{Z_i - z}{k_n}\right)}$$

Remarque :

- De manière abusive, les noyaux intervenant dans ces expressions sont tous représentés par la même lettre K, alors qu'ils sont fondamentalement différents, pour des raisons de dimension notamment.
- Les fenêtres sont, elles aussi, distinctes, h_n est la fenêtre associée à l'estimateur $\widehat{f_n}(x)$, et k_n celle associée à $\widehat{g_n}(z)$. Les vitesses de convergence de ces fenêtres seront ajustées à la dimension des régresseurs X et Z respectivement.
- Nous avons construit l'estimateur $\widehat{r_n}$ conformément aux estimateurs $\widehat{f_n}$ et $\widehat{g_n}$, utilisant les fenêtres h_n et k_n, en vue de simplifier sa définition. Cette construction est purement *ad hoc* et pourrait être améliorée.

Nous supposerons que les conditions sur les vitesses de convergence des fenêtres données par l'hypothèse 3.3 seront vérifiées pour chacune de ces deux fenêtres, ce que nous poserons sous la forme de l'hypothèse suivante :

Hypothèse 4.3 *(Conditions minimales sur les fenêtres) :*

Les fenêtres h_n et k_n vérifient les conditions de convergence

$$\lim_{n\to\infty} h_n = 0 \quad et \quad \lim_{n\to\infty} n \cdot h_n^p = \infty$$

et

$$\lim_{n\to\infty} k_n = 0 \quad et \quad \lim_{n\to\infty} n \cdot k_n^q = \infty$$

Ces conditions permettent aux estimateurs $\widehat{f_n}$, $\widehat{g_n}$ et $\widehat{r_n}$ d'être convergents dans leurs modèles respectifs (voir section 3.4).

Nous pouvons énoncer les premiers résultats de convergence de ces estimateurs.

Théorème 4.1 *Sous $\mathcal{H}_1$ et sous l'hypothèse 4.3, on a :*

$$i) \quad \widehat{f_n}(x) \xrightarrow{n\to\infty} f(x) \qquad \forall x$$

$$ii) \quad \widehat{g_n}(z) \xrightarrow{n\to\infty} E\left[f(x) \mid Z = z\right] \qquad \forall z$$

$$iii) \quad \widehat{r_n}(x,z) \xrightarrow{n\to\infty} f(x) \qquad \forall\ (x,z)$$

et

$$iv) \quad \widehat{\gamma} \xrightarrow{n\to\infty} \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot f\right]$$

Preuve[3] **:**

Ce résultat découle directement du corollaire 3.2, qui nous assure de la convergence en probabilité de l'estimateur du noyau de convolution. Sous l'hypothèse 4.3, on a

$$\begin{aligned}
\widehat{f_n}(x) &\xrightarrow{p} E[Y \mid X = x] = f(x) \\
\widehat{g_n}(z) &\xrightarrow{p} E[Y \mid Z = z] = g(z) \\
\widehat{r_n}(x,z) &\xrightarrow{p} E[Y \mid X = x,\, Z = z] = r(x,z)
\end{aligned}$$

Or sous l'hypothèse $\mathcal{H}_1$:

$$r(x,z) = E[Y \mid X = x,\, Z = z] = E[Y \mid X = x] = f(x)$$

et

$$\begin{aligned}
E[Y \mid Z = z] &= E[E[Y \mid X = x, Z = z] \mid Z = z] \\
&= E[E[Y \mid X = x] \mid Z = z] \\
&= E[f(x) \mid Z = z]
\end{aligned}$$

De plus l'estimateur $\widehat{\gamma}$ vérifie :

$$\begin{aligned}
\widehat{\gamma} &\xrightarrow{n\to\infty} (E[ZZ'])^{-1} E[ZY] \\
&= (E[ZZ'])^{-1} E[Z \cdot E[Y \mid X]] \\
&= (E[ZZ'])^{-1} E[Z \cdot f]
\end{aligned}$$

Ce qui montre le dernier point. □

Un résultat semblable est obtenu si l'on considère l'hypothèse d'indépendance linéaire, pour l'exclusion de la variable Z du modèle $\mathcal{M}_1$.

Théorème 4.2 *Sous les hypothèses* $(\mathcal{H}_2, \mathcal{H}_3)$*, on a :*

$$\begin{aligned}
&i) \quad \widehat{\beta} \xrightarrow{n\to\infty} \beta \\
&ii) \quad \widehat{\gamma} \xrightarrow{n\to\infty} (E[ZZ'])^{-1} E[Z \cdot X'] \cdot \beta \\
&iii) \quad \widehat{\alpha} \xrightarrow{n\to\infty} (\beta',\ 0)
\end{aligned}$$

[3]Les égalités intervenant dans les preuves des théorèmes 4.1 et 4.2 sont des égalités presque-sûres.

Si de plus l'hypothèse 4.3 est vérifiée, alors :

$$iv) \qquad \widehat{g_n}(z) \overset{n\to\infty}{\longrightarrow} \beta' \cdot E\left[X \mid Z = z\right] \quad \forall z$$

Preuve :
L'estimateur $\widehat{\beta}$ vérifie clairement :

$$\begin{aligned} \widehat{\beta} &\overset{n\to\infty}{\longrightarrow} \left(E\left[XX'\right]\right)^{-1} E\left[X \cdot Y'\right] \\ &= \beta \end{aligned}$$

Tandis que sous $\mathcal{H}_2$ et $\mathcal{H}_3$ l'estimateur $\widehat{\gamma}$ voit son comportement asymptotique modifié puisque :

$$\begin{aligned} \widehat{\gamma} &\overset{n\to\infty}{\longrightarrow} \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot Y'\right] \\ &\left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot L\left[Y \mid X, Z\right]'\right] \end{aligned}$$

Sous l'hypothèse $\mathcal{H}_2$, $L\left[Y \mid X, Z\right] = L\left[Y \mid X\right]$, d'ou :

$$\begin{aligned} \widehat{\gamma} &\overset{n\to\infty}{\longrightarrow} \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot L\left[Y \mid X, Z\right]'\right] \\ &= \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot L\left[Y \mid X\right]'\right] \\ &= \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot X'\right] \cdot \beta \end{aligned}$$

L'hypothèse 4.3 nous assure de la consistance de l'estimateur $\widehat{g_n}(z)$ nous avons donc :

$$\begin{aligned} \widehat{g_n}(z) &\overset{n\to\infty}{\longrightarrow} E\left[Y \mid Z = z\right] && \forall z \\ &= E\left[E\left[Y \mid X, Z\right] \mid Z = z\right] && \forall z \end{aligned}$$

L'hypohèse de linéarité $\mathcal{H}_3$ nous donne :

$$E\left[E\left[Y \mid X, Z\right] \mid Z = z\right] = E\left[L\left[Y \mid X, Z\right] \mid Z = z\right]$$

d'ou, sous l'hypothèse $\mathcal{H}_2$ maintenue

$$\begin{aligned} \widehat{g_n}(z) &\overset{n\to\infty}{\longrightarrow} E\left[L\left[Y \mid X, Z\right] \mid Z = z\right] && \forall z \\ &= E\left[L\left[Y \mid X\right] \mid Z = z\right] && \forall z \\ &= E\left[\beta' X \mid Z = z\right] && \forall z \end{aligned}$$

Ce qui montre le dernier point. □

Les limites des estimateurs $\widehat{\beta}$ et $\widehat{f_n}$ sous $\mathcal{M}_1$ ne dépendent pas de la distribution sous-jacente $\mathcal{P}_0$ des variables conditionnantes. Par contre les limites sous $\mathcal{M}_1$ des estimateurs du modèle $\mathcal{M}_2$ (qui est mal-spécifié pour $\mathcal{M}_1$) dépendent crucialement de cette distribution.

Cette dépendance disparaît lorsque le modèle rival emboîte le modèle $\mathcal{M}_1$, ce qui est le cas de $\mathcal{M}$. En effet les points (*iii*) des théorèmes 4.1 et 4.2 donnent les pseudo vraies valeurs associées à $\widehat{r_n}$ et à $\widehat{\alpha}$ indépendamment de la distribution des variables conditionnantes X et Z.

Ces résultats nous permettent de définir les pseudo-vraies valeurs associées aux estimateurs $\widehat{\gamma}$ et $\widehat{g_n}$.

4.2.3 Pseudo-vraies valeurs

Aux quatre situations possibles correspondent quatre pseudo-vraies valeurs. Conformément à la définition donnée section 1.7, les pseudo-vraies valeurs sont définies à partir des *plim* des estimateurs associés au modèle $\mathcal{M}_2$ données par les théorèmes 4.1 et 4.2. Ces pseudo- vraies valeurs lient les *"espaces paramétriques"* attachés aux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$. Ce terme est ici pris au sens large puisque ces *"espaces paramétriques"* pourront être fonctionnels. Les espaces Θ_f et Θ_g représenterons les espaces de fonctions associés aux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ lorsque ceux ci sont estimés non-paramétriquement, tandis que $\Theta_\beta \subset \Re^p$ et $\Theta_\gamma \subset \Re^q$ représenterons les espaces associés aux estimateurs $\widehat{\beta}$ et $\widehat{\gamma}$ respectivement.

Définition 4.2 *(Pseudo-vraies valeurs sous $\mathcal{H}_1$) :*

La pseudo-vraie valeur Γ associée à l'estimateur $\widehat{\gamma}$ sous $\mathcal{H}_1$ est :

$$\begin{array}{rccl} \Gamma & : & \Theta_f & \longrightarrow & \Theta_\gamma \\ & & f & \longrightarrow & \Gamma(f) = \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot f(X)\right] \end{array}$$

De même on définit la pseudo-vraie valeur G associée à l'estimateur $\widehat{g_n}$ sous $\mathcal{H}_1$ par :

$$\begin{array}{rccl} G & : & \Theta_f & \longrightarrow & \Theta_g \\ & & f & \longrightarrow & G(f)(z) = E\left[f(X) \mid Z = z\right] \end{array}$$

Il faut noter que lorsque l'estimateur paramétrique $\widehat{\gamma}$ est associé à $\mathcal{M}_2$, la pseudo-vraie valeur $\Gamma(f)$ est elle même paramétrique, elle sera d'ailleurs estimée de manière classique par $\widehat{\Gamma}(f)$. Dans le cadre non-paramétrique $G(f)(z)$

est une fonction de la variable z qui sera estimée par la même méthode que $g(z)$.

Deux autres pseudo-vraies valeurs sont également déduites du théorème 4.2, dans le cas où le modèle $\mathcal{M}_1$ est linéaire, conformément à l'hypothèse $\mathcal{H}_2$.

Définition 4.3 *(Pseudo-vraies valeurs sous $\mathcal{H}_2$) :*

La pseudo-vraie valeur Γ_L associée à l'estimateur $\widehat{\gamma}$ sous $\mathcal{H}_2$ est :

$$\begin{aligned} \Gamma_L \ : \quad \Theta_\beta &\longrightarrow \Theta_\gamma \\ \beta &\longrightarrow \Gamma_L(\beta) = \left(E\left[ZZ'\right]\right)^{-1} E\left[Z \cdot X'\right] \cdot \beta \end{aligned}$$

De même on définit la pseudo-vraie valeur G_L associée à l'estimateur $\widehat{g_n}$ sous $(\mathcal{H}_2\,, \mathcal{H}_3)$ par :

$$\begin{aligned} G_L \ : \quad \Theta_\beta &\longrightarrow \Theta_g \\ \beta &\longrightarrow G_L(\beta)(z) = \beta' \cdot E\left[X \mid Z = z\right] \end{aligned}$$

La nature des pseudo-vraies valeurs est conditionnée par la nature du paramètre associé à $\mathcal{M}_2$. Les pseudo-vraies valeurs G et G_L sont ainsi à valeur dans l'espace fonctionnel Θ_g, tandis que Γ_L et G_L sont à valeur dans $\Theta_\gamma \subset \Re^q$. Ces fonctions sont toutes linéaires en leurs arguments, que ceux-ci soient des vecteurs ou des fonctions. Elle peuvent s'interpréter également comme des projections entre espaces de vecteurs ou de fonctions.

Les pseudo-vraies valeurs introduites dans les définitions 4.2 et 4.3 sont théoriques puisque dépendantes du processus $\mathcal{P}_0$, elles doivent donc être estimées.

Les estimateurs paramétriques et non-paramétriques suivants seront utilisés pour l'estimation des pseudo-vraies valeurs.

Définition 4.4 *(Estimation des pseudo-vraies valeurs) :*

$$\begin{aligned} i) \qquad \widehat{\Gamma}(f) &= \left(\textstyle\sum_i Z_i Z_i'\right)^{-1} \sum_i Z_i \cdot f(X) \\ ii) \qquad \widehat{G}(f)(z) &= \frac{\sum_i f(X_i) \cdot K\left(\frac{Z_i - z}{k_n}\right)}{\sum_i K\left(\frac{Z_i - z}{k_n}\right)} \\ iii) \qquad \widehat{\Gamma_L}(\beta) &= \left(\textstyle\sum_i Z_i Z_i'\right)^{-1} \sum_i \left(Z_i X_i'\right) \beta \\ iv) \qquad \widehat{G_L}(f)(z) &= \frac{\sum_i X_i' \, \beta \cdot K\left(\frac{Z_i - z}{k_n}\right)}{\sum_i K\left(\frac{Z_i - z}{k_n}\right)} \end{aligned}$$

En appliquant les mêmes arguments que ceux utilisés dans les théorèmes 4.1 et 4.2, on montre que ces estimateurs $\widehat{\Gamma}(f)$, $\widehat{G}(f)$, $\widehat{\Gamma_L}$ et $\widehat{G_L}(f)$ sont des estimateurs convergents de $\Gamma(f)$, $G(f)$, Γ_L et $G_L(f)$ respectivement.

Remarque :

Nous avons considéré l'estimation sans imposer aucune contrainte sur le processus $\mathcal{P}_0$. Il existe toutefois des situations pour lesquelles ce processus est contraint. Bien que ces situations dépassent le cadre de notre étude nous discutons brièvement de telles situations sur deux exemples.

Une première restriction abordée en introduction, concerne la présence de variables communes aux deux modèles. Soit par exemple $X_i = (X_i^*, \xi_i)$ et $Z_i = (Z_i^*, \xi_i)$. Dans ce cas les définitions 4.2 et 4.3 des pseudo-vraies valeurs restent valides, en particulier la pseudo-vraie valeur G de la définition 4.2 s'écrit :

$$G(f)(z) = \int f(X)\varphi(X^* \mid Z, \xi)\, dX$$

et est estimée de manière consistante par $\widehat{G}(f)$ conformément à la définition 4.4.

Un deuxième exemple est tiré de Govaert et alii [43], dans le cadre dynamique où deux modèles autorégressifs sont proposés :

$$\begin{array}{rcl} & \mathcal{M}_1 \;\; : & \;\; Y_i = f(Y_{i-1}) + u_i \\ et & & \\ & \mathcal{M}_2 \;\; : & \;\; Y_i = g(Y_{i-2}) + v_i \end{array}$$

où : $f(Y_{i-1}) = E\left[Y_i \mid Y_0, \ldots, Y_{i-1}\right]$
Un estimateur non-paramétrique de g est :

$$\widehat{g}(y) = \frac{\sum_i Y_i \cdot K\left(\frac{Y_{i-2}-y}{k_n}\right)}{\sum_i K\left(\frac{Y_{i-2}-y}{k_n}\right)}$$

Si le processus est ergodique alors $\widehat{g}$ converge vers la pseudo-vraie valeur :

$$G(f) = \int f\left(Y_{i-1}\right) \varphi\left(Y_{i-1} \mid Y_{i-2}\right) dY_{i-1} = E\left[Y_i \mid Y_{i-2}\right]$$

Le calcul d'une matrice de covariance asymptotique pour la statistique définie sur la base de la différence $\left(\widehat{g}(y) - \widehat{G}(f)(y)\right)$ (telles que les statistiques définies ci-dessous), doit tenir compte des restrictions que comporte le processus $\mathcal{P}_0$. Nous ne discuterons pas davantage ces cas qui compliquent l'étude, laissant le lecteur intéressé se reporter à Govaert et alii pour une discussion générale sur l'enveloppement dans un contexte dynamique.

4.3 Statistiques d'enveloppement

Nous proposons de définir les différentes statistiques d'enveloppement de $\mathcal{M}_2$ par $\mathcal{M}_1$, en considérant une spécification paramétrique ou non-paramétrique pour chacun des deux modèles.

Dans chacune des quatre situations décrites par la table 4.1, la procédure de test d'enveloppement que nous construirons sera la même : nous évaluerons asymptotiquement la différence entre un estimateur de $\mathcal{M}_2$ et un estimateur de la pseudo-vraie valeur.

Il s'agira en fait de la différence entre *deux* estimateurs de $\mathcal{M}_2$, l'un réalisant l'estimation "*conventionelle*" de $\mathcal{M}_2$ (paramétrique ou non-paramétrique), l'autre estimant $\mathcal{M}_2$ dans la croyance que $\mathcal{M}_1$ est le "vrai" modèle. Cette différence, une fois normalisée, converge dans tous les cas vers une loi normale centrée, de laquelle nous tirerons une statistique distribuée asymptotiquement suivant une combinaison linéaire de lois χ^2.

Nous rappellerons tout d'abord les résultats paramétriques (PP) énoncés chapitre 2 et obtenus par Mizon et Richard [66], puis nous examinerons le cas complètement non-paramétrique (NN) enfin deux cas "*mixtes*" nous permettront de confronter modèles paramétriques et non-paramétriques (cas PN et NP). Les preuves complètes de chacun des résultats sont proposées en annexe, les principes de ces démonstrations seront toutefois exposés à la fin de chaque théorème.

Dans cette section nous supposerons l'homoscédasticité des résidus soit,

Hypothèse 4.4 *(Homoscédasticité des résidus) :*

Sous $\mathcal{M}_1$, $Var[Y \mid X, Z] = \sigma^2$, *inconnue.*

4.3.1 Enveloppement paramétrique (PP)

$\mathcal{M}_1$ et $\mathcal{M}_2$ sont deux modèles linéaires basés sur les X_i et Z_i respectivement. La statistique d'enveloppement est basée sur la différence entre $\widehat{\gamma}$ et $\widehat{\Gamma_L}(\widehat{\beta})$ l'estimateur de la pseudo-vraie valeur en $\widehat{\beta}$, soit :

$$\delta_{\beta,\gamma} = \widehat{\gamma} - \widehat{\Gamma_L}(\widehat{\beta})$$

Nous avons le résultat suivant :

Théorème 4.3 *: Sous* $\mathcal{H}_2$ *et* $\mathcal{H}_3$, *et sous l'hypothèse 4.4 :*

$$i) \qquad \sqrt{n} \cdot \delta_{\beta,\gamma} \quad \mathcal{D} \longrightarrow \quad \mathcal{N}\left(0, \sigma^2 \cdot V\left(\delta\right)\right)$$

$$ii) \qquad n \cdot \left(\delta'_{\beta,\gamma} \frac{\widehat{V(\delta)}^{+}}{\widehat{\sigma^2}} \delta_{\beta,\gamma}\right) \quad \mathcal{D} \longrightarrow \quad \chi^2_r$$

où :

$V(\delta) = Var(Z)^{-1} \cdot E\left[Var\left(Z \mid X\right)\right] \cdot Var(Z)^{-1}$

$\widehat{V(\delta)}^{+}$ *est un inverse généralisé d'un estimateur de la matrice* $V(\delta)$

r *est le rang de la matrice* $V(\delta)$

$\widehat{\sigma}^2$ *est un estimateur de* σ^2

Principe de la démonstration :

Nous décomposerons en premier lieu $\delta_{\beta,\gamma}$ de manière à faire apparaître le résidu $(Y_i - X_i'\widehat{\beta})$:

$$\delta_{\beta,\gamma} = \frac{\sum_i Z_i \left(Y_i - X_i'\widehat{\beta}\right)}{\sum_i Z_i Z_i'}$$

En introduisant β dans la décomposition de l'estimateur $\widehat{\beta}$, nous obtenons au numérateur une somme de termes indépendants centrés et de variance $\sigma^2 \cdot E\left[Var\left(Z \mid X\right)\right]$. Le théorème central limite nous donne alors le résultat □

4.3.2 Enveloppement non-paramétrique (NN)

Laissons cette fois les deux modèles libres de toute forme fonctionnelle, et considérons les modèles de régression non-paramétrique $\mathcal{M}_1$ et $\mathcal{M}_2$. Dans ce cas nous devons introduire des hypothèses supplémentaires sur le processus $\mathcal{P}_0$, sur la structure des noyaux et sur les fenêtres h_n et k_n.

Hypothèse 4.5 *(Existence et régularité de* f, φ *et de leurs dérivées d'ordre* d*) :*

- *Les densités et espérances conditionnelles issues de* (X, Y, Z) *sont* d*-fois continûment différentiables et bornées.*
- *Les densités marginales* $\varphi(\cdot)$ *sont à support compact et sont strictement positives sur ce support.*

Hypothèse 4.6 *(Régularité des noyaux) :*

Les noyaux intervenant dans les estimateurs $\widehat{f_n}$ *et* $\widehat{g_n}$ *sont de Parzen-Rosenblatt d'ordre* d*. Afin de travailler avec des noyaux positifs nous poserons* $d = 2$*.*

Hypothèse 4.7 *Les fenêtres* h_n *et* k_n *vérifient les conditions :*

$$n \cdot h_n^{p+2d} \overset{n\to\infty}{\longrightarrow} 0$$

et

$$n \cdot k_n^{q+2d} \overset{n\to\infty}{\longrightarrow} 0$$

Ces conditions[4] *nous permettent de tuer le biais asymptotique résultant de l'estimation par noyau de la fonction de régression, conformément aux résultats donnés section 3.4.*

A ces trois hypothèses nous assurant de la normalité asymptotique de $\widehat{f_n}$ et $\widehat{g_n}$ nous devons ajouter une hypothèse sur le rapport des fenêtres intervenant dans ces estimateurs.

Hypothèse 4.8 *Soit h_n et k_n les fenêtres associées aux estimateurs $\widehat{f_n}$ et $\widehat{g_n}$ respectivement. Ces fenêtres doivent vérifier :*

$$\log(n) \cdot \frac{k_n^q}{h_n^p} \stackrel{n\to\infty}{\longrightarrow} 0$$

Dans le cas univarié (p=q=1), cette hypothèse signifie intuitivement que la fenêtre k_n doit converger vers 0 "plus rapidement" que la fenêtre h_n.

Remarque :

Afin de mieux visualiser les contraintes imposées aux fenêtres h_n et k_n nous pouvons poser :

$$\begin{array}{rl} & h_n = n^{-a} \\ et & \\ & k_n = n^{-b} \end{array}$$

Les hypothèses 4.3 et 4.7 nous donnent :

$$\begin{array}{rl} & \frac{1}{p+2d} < a < \frac{1}{p} \\ et & \\ & \frac{1}{q+2d} < b < \frac{1}{q} \end{array}$$

L'hypothèse 4.8 ci-dessus impose la relation :

$$pa < qb$$

□

La statistique d'enveloppement que nous construisons dans ce cadre non-paramétrique est basée sur la fonction $\delta_{f,g}(z)$

$$\delta_{f,g}(z) = \widehat{g_n}(z) - \widehat{G}(\widehat{f})(z)$$

représentant la différence entre un estimateur de la fonction de régression de $\mathcal{M}_2$ et un estimateur de la pseudo-vraie valeur sous $\mathcal{M}_1$.

Sous les hypothèses précédentes nous pouvons énoncer la version non-paramétrique du théorème 4.3, concernant la statistique $\delta_{f,g}(\cdot)$.

[4] Ceci n'est d'ailleurs qu'une reformulation de l'hypothèse 3.4 (bis).

Théorème 4.4 *Sous $\mathcal{H}_1$ et sous les hypothèses 4.4, 4.5, 4.6,et si les fenêtres h_n et k_n vérifient les hypothèses 4.7 et 4.8, on a :*

$$\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) \xrightarrow{\mathcal{D}}$$

$$\longrightarrow \mathcal{N}\left(0, \frac{\sigma^2 \cdot \int K^2}{\varphi(z)}\right)$$

Principe de la démonstration :

De manière classique nous allons décomposer la statistique $\delta_{f,g}(z)$ afin de faire apparaître la différence$\left(Y_i - \widehat{f_n}(X_i)\right)$:

$$\delta_{f,g}(z) = \frac{\sum_i \left(Y_i - \widehat{f_n}(X_i)\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)}$$

En décomposant l'estimateur $\widehat{f_n}$, au point X_i en la somme $\widehat{f_n}(X_i) = f(X_i) + \left(\widehat{f_n}(X_i) - f(X_i)\right)$, nous obtenons :

$$\begin{aligned}
\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) &= \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i (Y_i - f(X_i)) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&\quad + \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(\widehat{f_n}(X_i) - f(X_i)\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&= A + B
\end{aligned}$$

Nous étudions ensuite le comportement asymptotique de ces deux termes sous nos hypothèses et nous montrons que :

$$A \xrightarrow{\mathcal{D}}$$

$$\longrightarrow \mathcal{N}\left(0, \frac{\sigma^2 \cdot \int K^2}{\varphi(z)}\right)$$

Tandis que :

$$B \xrightarrow{\mathcal{P}}$$

$$\longrightarrow 0$$

ce qui nous donne le résultat. □

Les cas "*mixtes*" permettant d'évaluer les capacités d'enveloppement d'un modèle non-paramétrique par un modèle paramétrique (et inversement) sont maintenant proposés.

4.3.3 Les cas "mixtes" (PN) et (NP)

Dans le premier, et le plus simple, des deux cas proposés ici, nous analysons le comportement de la statistique d'enveloppement basée sur un modèle $\mathcal{M}_1$ linéaire tandis que $\mathcal{M}_2$ est spécifié et estimé non-paramétriquement.

Enveloppement Paramétrique d'un modèle Non-paramétrique (PN)

On peut penser que l'enveloppement ne sera pas souvent vérifié ; toutefois s'il se trouve réalisé le résultat est très puissant. Cela signifie en effet, qu'un modèle linéaire basé sur les X_i explique l'ensemble des résultats d'un modèle $\mathcal{M}_2$ basé sur les Z_i, même si ce dernier est très général.

La statistique d'enveloppement est ici basée sur la fonction :

$$\delta_{\beta,g}(z) = \widehat{g_n}(z) - \widehat{G_L}(\widehat{\beta})(z)$$

Théorème 4.5 *(Cas P.N.)*
Sous $\mathcal{H}_2$, $\mathcal{H}_3$ et sous les hypothèses 4.4, 4.5, 4.6, et si la fenêtre k_n vérifie l'hypothèse 4.7, on a :

$$\sqrt{n \cdot k_n^q} \cdot \delta_{\beta,g}(z) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^2 \cdot \int K^2}{\varphi(z)}\right)$$

Principe de la démonstration :

Une nouvelle fois, la démonstration suit un schéma classique ; nous décomposons la statistique $\delta_{\beta,g}(z)$ pour faire apparaître le résidu $Y_i - X_i'\widehat{\beta}$:

$$\delta_{f,g}(z) = \frac{\sum_i \left(Y_i - X_i'\widehat{\beta}\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)}$$

L'introduction de $X_i'\beta$ dans ce terme nous permet de scinder le calcul en un terme C donnant la normalité asymptotiquement, et un terme D qui converge vers 0 :

$$\begin{aligned}
\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) &= \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(Y_i - X_i'\beta\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&\quad + \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(X_i'\widehat{\beta} - X_i'\beta\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&= C + D
\end{aligned}$$

□

Le dernier cas que nous considérerons concerne l'enveloppement du modèle $\mathcal{M}_2$ linéaire par un modèle $\mathcal{M}_1$ estimé non-paramétriquement.

Enveloppement Non-paramétrique d'un modèle Paramétrique (NP)

Comme dans les autres cas, la statistique $\delta_{f,\gamma}$ est basée sur une différence entre estimateurs de $\mathcal{M}_2$. Il s'agit ici de la différence entre un estimateur (paramétrique) de la régression et un estimateur (non-paramétrique) de la pseudo-vraie valeur sous $\mathcal{M}_1$:

$$\delta_{f,\gamma} = \widehat{\gamma} - \widehat{\Gamma}(\widehat{f_n})$$

Nous aurons besoin d'une hypothèse supplémentaire que nous substituerons à l'hypothèse 4.8 pour le théorème suivant :

Hypothèse 4.9 *La fenêtre h_n vérifie la condition :*

$$\sqrt{n} \cdot Max\left(\frac{\log(n)}{n \cdot h_n^p}\ ,\ h_n^{2d}\right) \overset{n\to\infty}{\longrightarrow} 0$$

Cette hypothèse sera satisfaite si la fenêtre $h_n = n^{-a}$ vérifie :

$$\frac{1}{4d} \leq a < \frac{1}{2p}$$

ce qui impose une relation entre d le degré de différentiabilité supposé des fonctions et la dimension de X, soit :

$$d \geq \frac{p}{2}$$

Dans le cas univarié nous devrons donc supposer que nos fonctions sont 2-fois continûment différentiables.

Nous pouvons énoncer le second théorème "*mixte*" :

Théorème 4.6 *: Sous $\mathcal{H}_1$ et sous les hypothèses 4.4, 4.5, 4.6, et si la fenêtre h_n vérifie les hypothèses 4.7 et 4.9 on a :*

$$i) \qquad \sqrt{n} \cdot \delta_{f,\gamma} \quad \mathcal{D} \longrightarrow \quad \mathcal{N}\left(0, \sigma^2 \cdot V(\delta)\right)$$

$$ii) \qquad n \cdot \delta'_{\beta,\gamma} \cdot \frac{\widehat{V(\delta)}^+}{\widehat{\sigma^2}} \cdot \delta_{\beta,\gamma} \quad \mathcal{D} \longrightarrow \quad \chi^2_r$$

où :

$$V(\delta) = Var(Z)^{-1} \cdot E\left[Var\left(Z \mid X\right)\right] \cdot Var(Z)^{-1}$$

$\widehat{V(\delta)}^{+}$ *est un inverse généralisé d'un estimateur*[5] *de la matrice* $V(\delta)$

r *est le rang de la matrice* $V(\delta)$

$\widehat{\sigma}^2$ *est un estimateur de* σ^2

Principe de la démonstration :

Aprés avoir regroupé les termes composant $\delta_{f,\gamma}$, et fait apparaître $\left(Y_i - \widehat{f_n}(X_i)\right)$dans son expression, l'introduction de $f(X_i)$ nous donne :

$$\begin{aligned}\sqrt{n}\cdot\delta_{f,\gamma} &= \frac{\sqrt{n}}{n}\cdot\frac{\sum_i Z_i(Y_i - f(X_i))}{\frac{1}{n}\sum_i Z_i Z_i'} \\ &\quad - \frac{\sqrt{n}}{n}\cdot\frac{\sum_i Z_i\left(\widehat{f_n}(X_i) - f(X_i)\right)}{\frac{1}{n}\sum_i Z_i Z_i'} \\ &= E + F\end{aligned}$$

Dans une première étape nous développerons $\widehat{f_n}(X_i) - f(X_i)$ par application d'une formule de Taylor :

$$\widehat{f_n}(X_i) - f(X_i) = \frac{\widehat{\Phi}(X_i) - f(x_i)\cdot\widehat{\varphi}(X_i)}{\varphi(X_i)} + R_n(X_i)$$

où $\widehat{f_n}(X_i) = \frac{\widehat{\Phi}(X_i)}{\widehat{\varphi}(X_i)}$

On montre alors que $\frac{\sqrt{n}}{n}\sum R_n(X_i)\cdot Z_i \xrightarrow{\mathcal{P}} 0$

La normalité asymptotique vient alors du terme E auquel s'ajoute la première partie de cette expression, il reste donc :

$$\left(\frac{1}{n}\sum_i Z_i Z_i'\right)\cdot G = \frac{\sqrt{n}}{n}\cdot\sum_i Z_i\left(Y_i - f(X_i)\right) - \frac{\sqrt{n}}{n}\cdot\sum_i \frac{\widehat{s}(X_i) - f(x_i)\cdot\widehat{\varphi}(X_i)}{\varphi(X_i)}$$

L'utilisation des définitions de $\widehat{\Phi}$ et de $\widehat{\varphi}$ suivie d'un traitement par U-statistiques nous permet d'écrire G sous la forme :

$$G = Var(Z)^{-1}\cdot\left[\frac{\sqrt{n}}{n}\sum_{i=1}^{n}\left(Y_i - m(X_i)\right)\left(Y_i - f(X_i)\right)\right] + O_p(1)$$

et de montrer ainsi que $G \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\cdot V(\delta)\right)$ □

[5] $\widehat{V(\delta)}$ peut être construit à partir des estimateurs non-paramétriques suivants :

$\widehat{Var(z)} = \frac{1}{n}\left(\sum_{i=1}^{n} Z_i\cdot Z_i' - \widehat{E}\left[Z \mid X = X_i\right]\cdot\widehat{E}\left[Z \mid X = X_i\right]'\right)$

et

$\widehat{E}\left[Z \mid X = X_i\right] = \frac{\sum_{j=1}^{n} Z_j\cdot K\left(\frac{X_j - X_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{X_j - X_i}{h_n}\right)}$

	Modèle M_2		
		Paramétrique	Non-paramétrique
Modèle M_1	Paramétrique	$\delta_{\beta,\gamma} = \widehat{\gamma} - \widehat{\Gamma_L}(\widehat{\beta})$	$\delta_{\beta,g} = \widehat{g} - \widehat{G_L}(\widehat{\beta})$
	Non-paramétrique	$\delta_{f,\gamma} = \widehat{\gamma} - \widehat{\Gamma}(\widehat{f})$	$\delta_{f,g} = \widehat{g} - \widehat{G}(\widehat{f})$

Table 4.2: Les quatre statistiques

4.4 Statistiques globales

Les comportements asymptotiques des statistiques reportées dans la table 4.2, présentent de nombreuses similitudes. Que ces statistiques soient réelles où fonctionnelles, la normalité asymptotique est obtenue dans tous les cas envisagés. Les lignes directrices des démonstrations sont les mêmes, et permettent d'identifier assez rapidement les éléments qui vont donner cette normalité, et ceux qui sont négligeables asymptotiquement. La situation se complique toutefois lorsque l'on cherche à obtenir une statistique de test globale à partir des statistiques fonctionnelles $\delta_{\beta,g}$ et $\delta_{f,g}$. Il nous a semblé utile, en effet, de chercher à obtenir une statistique globale obtenue à partir de ces statistiques "locales" (puisque fonctionnelles). Ces statistiques, semblables à celles déjà obtenues dans les cas PP et NP, peuvent êtres construites sur la base :

- d'un critère quadratique empirique

$$\Psi = \alpha_1(n) \cdot \sum_{i=1}^{n} \left(\delta_{\cdot,g}(Z_i)\right)^2 \varpi(Z_i)$$

- d'un critère type intégral

$$\Phi = \alpha_2(n) \cdot \int \left(\delta_{\cdot,g}(z)\right)^2 \varpi(z)\lambda(dz)$$

  ou de tout autre critère permettant une vision globale de la différence d'enveloppement

Les vitesses de convergence $\alpha_1(n)$ et $\alpha_2(n)$ sont alors choisies de manière à obtenir une convergence vers une distribution sur laquelle sera basée le test[6].

Avant de présenter section 4.4.3 le principal résultat obtenu sur ces statistiques de test, une réflexion sur les choix des fenêtres s'impose.

[6]Ces termes peuvent également faire intervenir "la", ou "les" fenêtres intervenant dans le calcul de $\delta_{f,g}$.

4.4.1 Fenêtres et stratégies

Nous avons insisté, dans le chapitre 3, sur l'importance du choix de la fenêtre sur la qualité des estimateurs, et nous avions volontairement relevé l'aspect arbitraire de ce choix. Un point concernant le cas NN mérite d'être souligné ; si le choix de la fenêtre dans l'un où l'autre des modèles n'est pas fait de manière automatique, des choix stratégiques peuvent influencer un test basé sur la statistique Ψ ou Φ.

En effet, la qualité des estimateurs dans chacun des modèles est déterminée par ces choix, mais ceux-ci peuvent devenir stratégiques dans une optique de comparaison ou de validation de modèles. Il peut paraître logique, à première vue, que le "propriétaire" du modèle $\mathcal{M}_1$ choisisse la fenêtre h dans le but d'estimer au mieux sa fonction de régression et que celui de $\mathcal{M}_2$ choisisse k dans le même esprit. Toutefois si aucune règle n'est imposée, un jeu stratégique peut s'instaurer entre des modélisateurs utilisant leurs fenêtres pour envelopper (ou ne pas se faire envelopper par) le concurrent.

Un autre point de vue se base sur le fait que $\mathcal{M}_1$ ne peut envelopper $\mathcal{M}_2$ pour n'importe quel choix de fenêtre. La notion d'enveloppement telle que nous l'avons présentée section 1.7 devrait donc incorporer explicitement la fenêtre, où une procédure d'estimation de celle-ci, comme un élément indissociable des estimateurs, et donc des modèles.

Enfin, nous avons volontairement occulté une fenêtre dans la première partie de ce chapitre : il s'agit de la fenêtre m intervenant dans l'estimation de la pseudo-vraie valeur $\widehat{G}(f)(z)$ (ou $\widehat{G_L}(\beta)(z)$ dans le cas PN). Cette fenêtre intervient dans le calcul de ces estimateurs et avait été arbitrairement posée comme égale à k, en toute logique la différence d'enveloppement $\delta_{f,g}(z)$ devrait s'écrire :

$$\begin{aligned}\delta_{f,g}(z) &= \widehat{g_k}(z) - \widehat{G_m}(\widehat{f_h})(z) \\ &= \frac{\sum_i Y_i K\left(\frac{Z_i - z}{k_n}\right)}{\sum_i K\left(\frac{Z_i - z}{k_n}\right)} - \frac{\sum_i \widehat{f_h}(X_i) K\left(\frac{Z_i - z}{m_n}\right)}{\sum_i K\left(\frac{Z_i - z}{m_n}\right)}\end{aligned}$$

et fait donc intervenir trois fenêtres :

- h associée à l'estimateur $\widehat{f_h}$
- k associée à l'estimateur $\widehat{g_k}$
- et m la "*pseudo-vraie fenêtre*" associée à l'estimateur $\widehat{G_m}$ de la pseudo vraie valeur G

Afin de mieux visualiser l'influence de chacune de ces fenêtres sur les statistiques d'enveloppement, nous proposons d'examiner, sur un exemple univarié,

Figure 4.1: $\Psi(h, k, m)$, m fixé

Figure 4.2: $\Psi(h, k, m)$, k fixé

la structure de la fonction $\Psi(h,k,m)$; comme il s'agit d'une fonction de trois paramètres nous présentons deux schémas différents[7] :

$\Psi(h,k,m)$ pour m fixé graphique 4.1 et $\Psi(h,k,m)$ pour k fixé, graphique 4.2.

Ces graphiques illustrent la forte influence que peut avoir un choix arbitraire des fenêtres sur la statistique Ψ. Sur notre exemple, le critère est croissant avec la fenêtre h et est décroissant avec k pour m fixé. Une stratégie simple pour le modélisateur de $\mathcal{M}_1$ désireux d'envelopper le modèle $\mathcal{M}_2$, (et donc désireux de minimiser Ψ), consiste à choisir une fenêtre h extrêmement petite et d'imposer une grande valeur pour k. En d'autres termes, il rend l'estimateur du modèle $\mathcal{M}_2$ peu informatif, de manière à pouvoir incorporer plus facilement ses résultats.

4.4.2 Sélection des fenêtres

L'automatisation des choix semble être la seule issue à ces conflits potentiels ; en supprimant tout arbitraire, on supprime toute prédominance abusive d'un modèle sur l'autre, et donc toute stratégie. Il convient toutefois de s'assurer que les critères sur lesquels sont basés la recherche des fenêtres sont judicieux. Une question se pose concernant la sélection de la pseudo-vraie fenêtre m : cette fenêtre doit-elle minimiser une procédure de "*bonne estimation*" de la pseudo-vraie valeur, ou doit elle minimiser le critère d'enveloppement $\Psi(h,k,m)$?

Nous pensons que la sélection de la pseudo-vraie fenêtre doit être indépendante du critère choisi pour mesurer l'enveloppement. Deux raisons motivent cette réponse :

- Le choix de m doit être antérieur au test, de même que le choix de l'estimateur de $G(f)(\cdot)$
- L'estimateur de $\widehat{G_m}$ doit converger vers la pseudo-vraie valeur G, ce qui n'est pas assuré si m minimise $\Psi(h,k,m)$.

Pour cela, nous nous proposons d'utiliser la technique de validation croisée, présentée section 3.5, pour la sélection de la fenêtre m ainsi que pour la sélection des fenêtres h et k.

Les fenêtres $\widehat{h}$ et $\widehat{k}$ seront donc définies par :

$$\widehat{h} = Arg \min_{h \in H_n} CV_x(h) \quad et \quad \widehat{k} = Arg \min_{k \in K_n} CV_z(k)$$

[7]Bien entendu cette fonction dépend des observations (X_i, Y_i, Z_i) et d'une manière plus générale du processus $\mathcal{P}_0$. Nous "*oublierons*" momentanément cette dépendance et concentrerons notre attention sur l'impact du trio (h,k,m) pour une observation. Le coefficient $\alpha_1(n)$ sera posé égal à 1.

où $CV(h)$

$$CV_{YX}(h) = \frac{1}{n}\sum_i \left(Y_i - \widehat{f_h^{-i}}(X_i)\right)^2 \cdot \varpi(X_i)$$

et

$$CV_{YZ}(k) = \frac{1}{n}\sum_i \left(Y_i - \widehat{g_k^{-i}}(Z_i)\right)^2 \cdot \varpi(Z_i)$$

la fenêtre $\widehat{m}$ est obtenue par minimisation sur l'intervalle M_n de :

$$CV_{fZ}(m) = \frac{1}{n}\sum_i \left(f(X_i) - \widehat{G_m^{-i}}(f)(Z_i)\right)^2 \cdot \varpi(Z_i)$$

Remarques :

- Le critère de validation croisée pour m ne pourra être effectué qu'après l'estimation de $f(X_i)$ par $\widehat{f_h}(X_i)$, on remplacera le critère $CV_f(m)$ par :

$$CV_{\widehat{f_h}Z}(m) = \frac{1}{n}\sum_i \left(\widehat{f_{\widehat{h}}}(X_i) - \widehat{G_m^{-i}}(\widehat{f_{\widehat{h}}})(Z_i)\right)^2 \cdot \varpi(Z_i)$$

  la fenêtre sélectionnée $\widehat{m}$, dépendra donc, de manière indirecte, de la fenêtre $\widehat{h}$.

- Dans le contexte PN, seules deux fenêtres, h et k sont à sélectionner puisque $\mathcal{M}_1$ est paramétrique. Dans le cas univarié, le critère de validation croisée pour la pseudo-vraie fenêtre m est alors :

$$CV_{\beta Z}(m) = \frac{1}{n}\sum_i \left(X_i.\beta - \widehat{G_m^{-i}}(\beta)(Z_i)\right)^2 \cdot \varpi(Z_i)$$

  ou, en utilisant la définition de l'estimateur leave-one-out $\widehat{G_L^{-i}}(\beta)(Z_i)$ de la pseudo-vraie valeur $\widehat{G_L}(\beta)$:

$$\begin{aligned}
CV_{\beta Z}(m) &= \tfrac{1}{n}\textstyle\sum_{i\neq j}\left(X_i.\beta - \dfrac{\sum_{j\neq i} X_j\beta\cdot K\left(\frac{Z_i-Z_j}{m}\right)}{\sum_i K\left(\frac{Z_i-Z_j}{m}\right)}\right)^2\cdot\varpi(Z_i)\\
&= \beta^2\cdot\tfrac{1}{n}\textstyle\sum_{i\neq j}\left(X_i - \dfrac{\sum_{j\neq i} X_j\cdot K\left(\frac{Z_i-Z_j}{m}\right)}{\sum_i K\left(\frac{Z_i-Z_j}{m}\right)}\right)^2\cdot\varpi(Z_i)\\
&= \beta^2\cdot\tfrac{1}{n}\textstyle\sum_{i\neq j}\left(X_i - \widehat{g_m^{-i}}(Z_i)\right)^2\cdot\varpi(Z_i)\\
&= \beta^2\cdot CV_{XZ}(m)
\end{aligned}$$

Nous trouvons ainsi, fort logiquement, le critère $CV_{XZ}(m)$ de validation croisée pour la détermination de la fenêtre intervenant dans l'estimation de la régression $r(\cdot) = E[X \mid Z = \cdot]$.

Les intervalles H_n et K_n donnés par Härdle et Marron [50] vont nous permettre d'imposer quelques *"garde-fous"* pour conserver nos hypothèses tout en automatisant la recherche des fenêtres. Ces intervalles sont de la forme :

$$H_n = \left[\underline{h}, \overline{h}\right] = \left[a_n \cdot n^{-\frac{1}{p}}\,,\, (a_n)^{-1}\right]$$

$$K_n = \left[\underline{k}, \overline{k}\right] = \left[b_n \cdot n^{-\frac{1}{q}}\,,\, (b_n)^{-1}\right]$$

où $a_n = n^{\delta_1}$ et $b_n = n^{\delta_2}$, δ_1 et δ_2 sont des constantes positives. Härdle ([46] p.159) nous indique qu'il est possible d'effectuer la recherche de cette fenêtre sur un intervalle du type $\left[c_n \cdot n^{-\frac{1}{p+2d}}\,,\, d_n \cdot n^{-\frac{1}{p+2d}}\right]$.

En effet, afin de nous assurer que les fenêtres $\widehat{h}$ et $\widehat{k}$ vérifient les hypothèses 4.7 nous pouvons insérer l'intervalle $H_n = \left[\underline{h}, \overline{h}\right]$ dans un intervalle du type $\left[c_n \cdot n^{-\frac{1}{p+2d}}\,,\, d_n \cdot n^{-\frac{1}{p+2d}}\right]$. Pour obtenir une telle situation on devra vérifier les conditions suivantes :

$$\underline{h} = a_n \cdot n^{-\frac{1}{p}} > c_n \cdot n^{-\frac{1}{p+2d}} \Leftrightarrow \delta_1 > \frac{2d}{p(2d+p)}$$

et

$$\overline{h} = (a_n)^{-1} < d_n \cdot n^{-\frac{1}{p+2d}} \Leftrightarrow \delta_1 > \frac{1}{2d+p}$$

Ainsi si $d_n \overset{n\to\infty}{\longrightarrow} 0$, la condition 4.7

$$n \cdot \widehat{h}^{p+2d} \longrightarrow 0$$

sera vérifiée pour la grille de sélection H_n. On opère de même pour la grille K_n.

L'hypothèse 4.8 concernant le rapport des fenêtres h_n et k_n peut également être incorporée par l'intermédiaires des intervalles H_n et K_n. Il faut pour cela imposer que :

$$\log(n) \cdot \frac{(\underline{k})^q}{\left(\overline{h}\right)^p} \overset{n\to\infty}{\longrightarrow} 0$$

c'est-à-dire :

$$q\delta_2 + p\delta_1 < 1$$

4.4.3 Statistique globale non-paramétrique

Nous présentons ici un résultat concernant une évaluation "globale" de la différence d'enveloppement entre deux modèles non-paramétriques univariés. Nous nous intéressons à :

$$\int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz = \int \left(\widehat{g_n}(z) - \widehat{G}(\widehat{f})(z)\right)^2 \varpi(z)dz$$

Contrairement aux cas paramétriques (PP) étudiés précédemment, nous n'obtenons une convergence vers une loi classique (normale centrée ou χ^2), mais vers ce que nous appellerons une loi normale "*fuyante*". En fait, nous obtenons que :

$$n\sqrt{k} \cdot \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz = B \cdot \frac{1}{\sqrt{k}} + n\sqrt{k} \cdot I_{n,2} + O_p\left(k\right)$$

où

$$n \cdot \sqrt{k} \cdot I_{n,2} \xrightarrow{\mathcal{D}}$$

$\longrightarrow \mathcal{N}\left(0, \alpha_1\right)$

L'analogie avec le cas paramétrique n'est donc pas vérifiée. Ce résultat n'est pas sans rappeler un résultat de Härdle et Mammen [49], dans le cadre d'une comparaison entre un modèle paramétrique et un modèle non-paramétrique. Nous discuterons des implications de ce résultat et de ses liens avec la littérature après avoir énoncé ce théorème.

Les hypothèses que nous formulons sont semblables à celle posées dans la section 4.2 pour la formulation du théorème 4.4 (cas NN).

Hypothèse 4.10 *(Existence et régularité de f, g, φ et de leurs dérivées d'ordre 2) :*

- *Les densités et espérances conditionnelles issues de (X, Y, Z) sont 2-fois continûment différentiables et bornées.*
- *Les densités marginales $\varphi(\cdot)$ sont à support compact et sont strictement positives sur ce support.*

Cette hypothèse est une reformulation de l'hypothèse 4.5, dans le cas d=2.

Hypothèse 4.11 *Le noyau K est une densité bornée sur $\Re$ à support compact*

$$\int uK(u)du = 0 \quad et \quad \int u^2K(u)du = C$$

Afin de simplifier nos calculs nous prendrons le même noyau pour $\widehat{f_n}$ et pour $\widehat{g_n}$, nos résultats restent toutefois valides pour des noyaux différents.

Hypothèse 4.12 *(Existence et régularité des moments conditionnels)*

$$f_4(x) = E\left[(Y - f(x))^4 \mid X = x\right]$$

est telle que

$$\int f_4(x)\varphi(x)dx < \infty$$

Hypothèse 4.13 *Les fenêtres h_n et k_n vérifient les hypothèses de convergence suivantes :*

$$nh_n^5 \overset{n\to\infty}{\longrightarrow} \infty$$

et

$$nk_n^5 \overset{n\to\infty}{\longrightarrow} \infty$$

Cette hypothèse correspond à l'hypothèse 4.7, dans le cadre univarié p=q=1 et d=2.

Hypothèse 4.14 *Nous imposerons la condition supplémentaire suivante :*

$$\log(n)\frac{\sqrt{k_n}}{h_n} \overset{n\to\infty}{\longrightarrow} 0$$

Théorème 4.7 *Sous $(\mathcal{H}_1, \mathcal{H}_4)$, sous les hypothèses 4.10, 4.11 et 4.12 et si les fenêtres h_n et k_n vérifient les hypothèses 4.13 et 4.14, on a :*

$$\begin{aligned} n \cdot \sqrt{k} \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz &= \tfrac{1}{\sqrt{k}}\, \sigma^2 \cdot \int K^2(u)\nu_n(u)du \\ &\quad + n\sqrt{k} \cdot I_{n,2} \\ &\quad + O_p(k) + O_p\left(\tfrac{1}{\sqrt{n\cdot k}}\right) \end{aligned}$$

Dans ces expressions nous avons :

- Le terme $n\sqrt{k}\ I_{n,2}$ converge asymptotiquement vers une loi normale centrée :

$$n \cdot \sqrt{k} \cdot I_{n,2} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, \alpha_1)$$

où :

$$\alpha_1 = 2 \cdot \sigma^4 \left[\int \varphi^2(x)\nu(x)dx\right] \cdot \left[\int \left[\int K(u)K(u+v)du\right]^2 dv\right]$$

- La fonction de poids $\varpi(z)$ a été choisie de la forme :

$$\varpi(z) = \widehat{\varphi_n}(z) \cdot \nu_n(z)$$

où $\nu_n(z)$ est telle qu'il existe une fonction $\nu(\cdot)$, bornée, positive telle que :

$$\sup_{x\in\Re} \left| \frac{\nu_n(z)}{\nu(z)} - 1 \right| \overset{P}{\longrightarrow} 0$$

Principe de la démonstration :

Cette démonstration s'inspire des travaux de Hall [45] et de Härdle et Mammen [49], nous décomposons tout d'abord le terme $\psi = \int \left(\delta_{f,g}(z)\right)^2 \varpi(z)dz$ en deux termes

$$\begin{aligned}\psi &= \tfrac{1}{n^2k^2} \int \left(\sum_i \left(Y_i - \widehat{f}(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right)\right)^2 \nu_n(z)dz \\ &= \tfrac{1}{n^2k^2} \int \left(\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right) + \sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right)^2 \nu_n(z)dz \\ &= \tfrac{1}{n^2k^2} \int \left(Q_1(z) + Q_2(z)\right)^2 \nu_n(z)dz\end{aligned}$$

En développant ce carré, en la somme des carrés et du double produit, nous obtenons trois termes qui seront étudiés séparément :

$$F_1 = \tfrac{1}{n^2k^2} \int \left[\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right)\right]^2 \nu_n(z)dz$$

$$F_2 = \tfrac{1}{n^2k^2} \int \left[\sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right]^2 \nu_n(z)dz$$

et

$$F_3 = \tfrac{1}{n^2k^2} \int \left[\sum_i \left(Y_i - f(X_i)\right) K\left(\tfrac{Z_i - z}{k}\right)\right] \cdot \left[\sum_j \left(f(X_j) - \widehat{f}(X_j)\right) K\left(\tfrac{Z_j - z}{k}\right)\right] \nu_n(z)dz$$

Nous montrons ensuite que **:**

- F_1 nous donne deux termes non nuls asymptotiquement :

$$\begin{aligned}F_1 &= \tfrac{1}{nk}\,\sigma^2 \cdot \int K^2(u)\nu_n(u)du \\ &\quad + I_{n,2} \\ &\quad + O_p\left(\tfrac{1}{\sqrt{n^3 \cdot k^2}}\right)\end{aligned}$$

où

$$\left(n \cdot \sqrt{k}\right) \cdot I_{n,2} \xrightarrow{\mathcal{D}}$$

$\longrightarrow \mathcal{N}\left(0, \frac{1}{4} \cdot \alpha_1\right)$

- $n\sqrt{k} \cdot F_2 \to 0$ et l'on peut s'assurer de son comportement par un bon choix de la fenêtre "h" qu'il fait intervenir, car :

$$F_2 \leq O_p\left(Max\left(\frac{\log(n)}{n \cdot h}\ ,\ h^{2 \cdot d}\right)\right) \cdot \int \widehat{\varphi}^2(z)\nu_n(z)dz$$

- $n\sqrt{k} \cdot F_3 \to 0$ assez lentement, on montre en effet que

$$n\sqrt{k} \cdot F_3 = O_p(k)$$

Comme

$$n\sqrt{k} \cdot \psi = n\sqrt{k} \cdot (F_1 + F_2 + 2 \cdot F_3)$$

nous obtenons le résultat. □

Remarque

Ce résultat est semblable à celui donné par Härdle et Mammen [49], dans le cadre d'une comparaison entre un modèle paramétrique *"lissé"* et un modèle non-paramétrique. Les deux modèles sont basés sur les mêmes ensembles de régresseurs, X_i mais deux estimateurs $f_{\widehat{\theta}}$ (paramétrique) et $\widehat{f}$ (non-paramétrique) sont construits sur la base de n observations *iid*, $(X_i, Y_i)_{i=1,\ldots,n}$, et l'on cherche à déterminer les différences qui peuvent exister entre ces deux estimateurs. La comparaison s'effectue sur la base d'un critère intégral semblable au nôtre :

$$T_n = n\sqrt{h^p} \int \left(\widehat{f}(x) - F(f_{\widehat{\theta}})(x)\right)^2 \varpi(x)dx$$

où $\varpi(x)$ est une fonction de poids et $F(f_{\widehat{\theta}})(x)$ est l'estimateur non-paramétrique de la régression paramétrique $f_\theta(x)$ pour $\theta = \widehat{\theta}$, soit :

$$F(f_{\widehat{\theta}})(x) = \frac{\sum_i f_{\widehat{\theta}}(X_i) \cdot K\left(\frac{X_i - x}{h}\right)}{\sum_i K\left(\frac{X_i - x}{h}\right)}$$

Un même comportement asymptotique de normale "*fuyante*" est alors observé. Une deuxième caractéristique commune à ces résultats est la vitesse de convergence en $n\sqrt{h}$ que nous obtenons pour les deux statistiques.

Même si le terme "fuyant" est estimable, l'approche asymptotique n'est donc pas la meilleure pour l'étude de notre statistique. Devant ce constat, nous pouvons nous joindre à Härdle et Mammen : *"Therefore the theorem can only give a rough idea of the stochastic behavior of T_n if the sample is small."*

Une solution proposée consiste à "*Bootstraper*" cette statistique afin d'en obtenir la distribution. Les méthodes de Bootstrap semblent en effet particulièrement adaptées à notre problème. Elles constitueraient ainsi une alternative à l'approche asymptotique.

4.5 Annexe au chapitre 4

Démonstration du théorème 4.3 (Cas PP)

On peut écrire la statistique $\delta_{\beta,\gamma}$ sous la forme :

$$\begin{aligned}\delta_{\beta,\gamma} &= \left(\widehat{\gamma} - \widehat{\Gamma_L}(\widehat{\beta})\right) \\ &= \left(\frac{\sum_i Z_i Y_i}{\sum_i Z_i Z_i^{'}} - \frac{\sum_i Z_i X_i^{'} \cdot \widehat{\beta}}{\sum_i Z_i Z_i^{'}}\right) \\ &= \left(\sum_i Z_i Z_i^{'}\right)^{-1} \left(\sum_i Z_i \left(Y_i - X_i^{'} \cdot \widehat{\beta}\right)\right)\end{aligned}$$

En décomposant l'estimateur $\widehat{\beta}$ en la somme $\widehat{\beta} = \beta + \left(\widehat{\beta} - \beta\right)$, soit :

$$\begin{aligned}\widehat{\beta} &= \beta + \tfrac{1}{\sum_i X_i X_i^{'}} \cdot \left(\sum_i X_i Y_i - \sum_i X_i X_i^{'} \cdot \beta\right) \\ &= \beta + \tfrac{1}{\sum_i X_i X_i^{'}} \cdot \left(\sum_i X_i \left(Y_i - X_i^{'}\beta\right)\right)\end{aligned}$$

D'où nous tirons :

$$\begin{aligned}\delta_{\beta,\gamma} &= \left(\sum_i Z_i Z_i^{'}\right)^{-1} \left(\sum_i Z_i Y_i - \sum_i Z_i X_i^{'} \cdot \beta - \frac{\sum_i Z_i X_i^{'}}{\sum_i X_i X_i^{'}} \cdot \left(\sum_i X_i \left(Y_i - X_i^{'}\beta\right)\right)\right) \\ &= \left(\sum_i Z_i Z_i^{'}\right)^{-1} \left(\sum_i Z_i \left(Y_i - X_i^{'}\beta\right) - \frac{\sum_j Z_j X_j^{'}}{\sum_j X_j X_j^{'}} \sum_i X_i \left(Y_i - X_i^{'}\beta\right)\right) \\ &= \left(\sum_i Z_i Z_i^{'}\right)^{-1} \left(\sum_i \left(Z_i - \frac{\sum_j Z_j X_j^{'}}{\sum_j X_j X_j^{'}} \cdot X_i\right) \left(Y_i - X_i^{'}\beta\right)\right)\end{aligned}$$

Sous les hypothèses $\mathcal{H}_2$ et $\mathcal{H}_3$, le terme général de cette somme vérifie :

$$E\left[\left(Z_i - \frac{\sum_j Z_j X_j^{'}}{\sum_j X_j X_j^{'}} \cdot X_i\right) \left(Y_i - X_i^{'}\beta\right)\right] = 0$$

sa variance est donc:

$$E\left[\left(Z_i - \frac{\sum_j Z_j X_j^{'}}{\sum_j X_j X_j^{'}} \cdot X_i\right)^2 \left(Y_i - X_i^{'}\beta\right)^2\right]$$

soit encore :

$$= E\left[E\left[\left(Z_i - \frac{\sum_j Z_j X_j'}{\sum_j X_j X_j'} \cdot X_i\right)^2 \cdot (Y_i - X_i'\beta)^2 \mid X_i, Z_i\right]\right]$$

en utilisant l'hypothèse d'homoscédasticité 4.4, il vient :

$$Var\left[\left(Z_i - \frac{\sum_j Z_j X_j'}{\sum_j X_j X_j'} \cdot X_i\right)(Y_i - X_i'\beta)\right] = \sigma^2 \cdot E\left[Var\left(Z \mid X\right)\right]$$

d'où le résultat par application du théorème central limite. □

Démonstration du théorème 4.4 (Cas NN)

D'une manière qui deviendra classique dans nos démonstrations nous allons décomposer la statistique $\delta_{f,g}(z)$ de manière à faire apparaître la différence[8] $\left(Y_i - \widehat{f_n}(X_i)\right)$:

$$\begin{aligned}
\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) &= \sqrt{n \cdot k_n^q} \cdot \left(\widehat{g_n}(z) - \widehat{G}(\widehat{f})(z)\right) \\
&= \sqrt{n \cdot k_n^q} \cdot \left(\frac{\sum_i Y_i K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} - \frac{\sum_i \widehat{f_n}(X_i)_i K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)}\right) \\
&= \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(Y_i - \widehat{f_n}(X_i)\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)}
\end{aligned}$$

En décomposant l'estimateur $\widehat{f_n}$, au point X_i en la somme $\widehat{f_n}(X_i) = f(X_i) + \left(\widehat{f_n}(X_i) - f(X_i)\right)$, nous obtenons :

$$\begin{aligned}
\sqrt{n \cdot k_n^q} \cdot \delta_{f,g}(z) &= \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i (Y_i - f(X_i)) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&\quad + \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(\widehat{f_n}(X_i) - f(X_i)\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\
&= A + B
\end{aligned}$$

A est l'estimateur du noyau de la régression du résidu $U_i = Y_i - f(X_i)$ sur les Z_i dont le comportement asymptotique sous $\mathcal{H}_1$ est donné par le théorème 3.5, soit :

$$A \ \mathcal{D}$$

[8] Dans la démonstration du théorème 4.3, nous faisions apparaître la différence $\left(Y_i - X_i' \cdot \widehat{\beta}\right)$.

$\longrightarrow \mathcal{N}\left(0, \frac{\sigma^2 \cdot \int K^2}{\varphi(z)}\right)$

Pour pouvoir conclure, il nous faut montrer la convergence vers 0 du second terme B, l'hypothèse 4.6, nous indique que le noyau K est positif, ce qui nous permet de majorer le terme B :

$$\begin{aligned}\sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(\widehat{f_n}(X_i) - f(X_i)\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} &\leq \sqrt{n \cdot k_n^q} \cdot \frac{\sum_i \left(\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right|\right) K\left(\frac{Z_i - z}{h_n}\right)}{\sum_i K\left(\frac{Z_i - z}{h_n}\right)} \\ &\leq \sqrt{n \cdot k_n^q} \cdot \left(\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right|\right)\end{aligned}$$

Le théorème 3.3 nous indique que la quantité $\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right|$ converge vers 0 et nous donne la clé pour conclure cette démonstration en nous indiquant la vitesse de convergence de $\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right|$.

En effet,

$$\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right| = O_p\left[\max\left(\frac{\sqrt{\log(n)}}{\sqrt{n \cdot h_n^p}}\ ,\ h_n^d\right)\right]$$

Notre terme B est donc d'ordre $\sqrt{n \cdot k_n^q} \cdot O_p\left[\max\left(\frac{\sqrt{\log(n)}}{\sqrt{n \cdot h_n^p}}\ ,\ h_n^d\right)\right]$, soit :

$$\sup_{X_i} \left|\widehat{f_n}(X_i) - f(X_i)\right| = O_p\left[\max\left(\frac{\sqrt{\log(n)} \cdot \sqrt{n \cdot k_n^q}}{\sqrt{n \cdot h_n^p}}\ ,\ \sqrt{n \cdot k_n^q} \cdot\ h_n^d\right)\right]$$

L'hypothèse 4.8, nous permet d'affirmer que $B \xrightarrow{\mathcal{P}} 0$ puisque :

$$\frac{\log(n) \cdot k_n^q}{h_n^p} \xrightarrow{n \to \infty} 0$$

et

$$\sqrt{n \cdot k_n^q} \cdot\ h_n^d \leq \sqrt{n \cdot h_n^p} \cdot\ h_n^d$$

La fenêtre h_n vérifiant l'hypothèse 4.7, on a $\sqrt{n \cdot h_n^p} \cdot\ h_n^d \xrightarrow{n \to \infty} 0$.

En utilisant le théorème de Slutsky (Serfling [79], p.19), nous obtenons que la somme $A + B \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\sigma^2 \cdot \int K^2}{\varphi(z)}\right)$, ce qui nous donne le résultat. □

Démonstration du théorème 4.5 (Cas PN)

Une nouvelle fois, notre premier travail consiste à décomposer la statistique $\delta_{f,g}(z)$ de manière à faire apparaître le résidu $\left(Y_i - X_i' \widehat{\beta}\right)$:

$$\begin{aligned}\sqrt{n\cdot k_n^q}\cdot\delta_{\beta,g}(z) &= \sqrt{n\cdot k_n^q}\cdot\left(\widehat{g_n}(z)-\widehat{G_L}(\widehat{\beta})(z)\right)\\ &= \sqrt{n\cdot k_n^q}\cdot\left(\frac{\sum_i Y_i K\left(\frac{Z_i-z}{h_n}\right)}{\sum_i K\left(\frac{Z_i-z}{h_n}\right)}-\frac{\sum_i X_i'\,\widehat{\beta}\cdot K\left(\frac{Z_i-z}{k_n}\right)}{\sum_i K\left(\frac{Z_i-z}{k_n}\right)}\right)\\ &= \sqrt{n\cdot k_n^q}\cdot\frac{\sum_i\left(Y_i-X_i'\,\widehat{\beta}\right)K\left(\frac{Z_i-z}{h_n}\right)}{\sum_i K\left(\frac{Z_i-z}{h_n}\right)}\end{aligned}$$

En décomposant $\widehat{\beta}$ en la somme $\widehat{\beta}=\beta+\left(\widehat{\beta}-\beta\right)$, nous obtenons :

$$\begin{aligned}\sqrt{n\cdot k_n^q}\cdot\delta_{f,g}(z) &= \sqrt{n\cdot k_n^q}\cdot\frac{\sum_i\left(Y_i-X_i'\,\beta\right)K\left(\frac{Z_i-z}{h_n}\right)}{\sum_i K\left(\frac{Z_i-z}{h_n}\right)}\\ &\quad+\sqrt{n\cdot k_n^q}\cdot\frac{\sum_i\left(X_i'\,\widehat{\beta}-X_i'\,\beta\right)K\left(\frac{Z_i-z}{h_n}\right)}{\sum_i K\left(\frac{Z_i-z}{h_n}\right)}\\ &= C+D\end{aligned}$$

Nous utiliserons la même technique que celle utilisée dans la démonstration du théorème 4.4, puisque C est l'estimateur du noyau de la régression du résidu $U_i=Y_i-f(X_i)$ sur les Z_i dont le comportement asymptotique sous $\mathcal{H}_2$ et $\mathcal{H}_3$ est donné par le théorème 3.5, soit :

$$C \xrightarrow{\mathcal{D}}$$

$$\longrightarrow \mathcal{N}\left(0,\frac{\sigma^2\cdot\int K^2}{\varphi(z)}\right)$$

Tandis que le terme $D \xrightarrow{\mathcal{P}}$
$\longrightarrow 0$, puisque :

$$\begin{aligned}D &= \sqrt{n\cdot k_n^q}\cdot\frac{\sum_i\left(X_i'\widehat{\beta}-X_i'\beta\right)\cdot K\left(\frac{Z_i-z}{h_n}\right)}{\sum_i K\left(\frac{Z_i-z}{h_n}\right)}\\ &= \left(\widehat{\beta}-\ \beta\right)\cdot\sqrt{n\cdot k_n^q}\,\widehat{E}\left[X'\mid Z=z\right]\\ &\xrightarrow{\mathcal{P}} 0\end{aligned}$$

d'où le résultat par application du théorème de Slutsky . □

Démonstration du théorème 4.6 (Cas NP)

Une nouvelle fois on peut décomposer la statistique $\delta_{f,\gamma}$ de manière à faire apparaître le résidu $\left(Y_i - \widehat{f_n}(X_i)\right)$ de la régression estimée sous $\mathcal{H}_1$:

$$\begin{aligned}\sqrt{n}\cdot\delta_{f,\gamma} &= \sqrt{n}\cdot\left(\widehat{\gamma} - \widehat{\Gamma}(\widehat{f_n})\right)\\ &= \sqrt{n}\cdot\left(\frac{\sum_i Z_i Y_i}{\sum_i Z_i Z_i'} - \frac{\sum_i Z_i\cdot\widehat{f}_n(X_i)}{\sum_i Z_i Z_i'}\right)\\ &= \sqrt{n}\cdot\frac{\sum_i Z_i\left(Y_i - \widehat{f}_n(X_i)\right)}{\sum_i Z_i Z_i'}\end{aligned}$$

La décomposition de ce résidu $\left(Y_i - \widehat{f_n}(X_i)\right) = (Y_i - f(X_i)) - \left(\widehat{f_n}(X_i) - f(X_i)\right)$, nous donne :

$$\begin{aligned}\sqrt{n}\cdot\delta_{f,\gamma} &= \frac{\sqrt{n}}{n}\cdot\frac{\sum_i Z_i(Y_i - f(X_i))}{\frac{1}{n}\sum_i Z_i Z_i'}\\ &\quad - \frac{\sqrt{n}}{n}\cdot\frac{\sum_i Z_i\left(\widehat{f}_n(X_i) - f(X_i)\right)}{\frac{1}{n}\sum_i Z_i Z_i'}\\ &= E - F\end{aligned}$$

Nous allons étudier séparément ces deux termes qui n'ont pas le même comportement asymptotique sous $\mathcal{H}_1$. Nous commencerons par traiter F, une partie de ce terme va s'ajouter a E pour former G et donner la normalité, tandis que le restant disparaît asymptotiquement.

En effet sous $\mathcal{H}_1$ et sous les hypothèses 4.4, 4.5, 4.6, et 4.7, on a :

Sous l'hypothèse supplémentaire 4.9

- $F = F_1 + F_2$ où $F_2 \xrightarrow{\mathcal{P}}$

$\longrightarrow 0$, tandis que

- $G = E + F_1 \xrightarrow{\mathcal{D}}$

$\longrightarrow \mathcal{N}\left(0, \sigma^2\cdot V(\delta)\right)$

Première étape :

La limite du dénominateur de F ne pose aucun problème, c'est $Var(Z)$, nous occuperons donc principalement du numérateur que nous notons F^+.

Nous pouvons réécrire la différence $\left(\widehat{f_n}(X_i) - f(X_i)\right)$ sous la forme :

$$\widehat{f_n}(X_i) - f(X_i) = \frac{\widehat{\Phi}(X_i)}{\widehat{\varphi}(X_i)} - \frac{\Phi(X_i)}{\varphi(X_i)}$$

où $\Phi(\cdot) = \int\varphi(\cdot, y)dy$ et $\widehat{\Phi}(\cdot) = \frac{1}{nh^p}\sum_i Y_i K\left(\frac{X_i - \cdot}{h}\right)$ est son estimateur non-paramétrique :

en utilisant le développement de la fonction :

$$p(x,y) \longrightarrow \frac{x}{y}$$

On a :

$$\begin{aligned} p(x,y) - p(x_1,y_1) &= (x-x_1)/y \\ &\quad -x/y^2 \cdot (y-y_1) \\ &\quad -(y-y_1)^2 \cdot x_\alpha / y_\alpha^3 \\ &\quad -\tfrac{1}{2}(x-x_1)(y-y_1)/y_\alpha \end{aligned}$$

où $x_\alpha \in [x, x_1]$ et $y_\alpha \in [y, y_1]$
on obtient ainsi :

$$\begin{aligned} \widehat{f_n}(X_i) - f(X_i) &= \tfrac{1}{\varphi(X_i)} \cdot \left(\widehat{\Phi}(X_i) - \Phi(X_i)\right) \\ &\quad -\tfrac{\Phi(X_i)}{\varphi(X_i)^2} \left(\widehat{\varphi}(X_i) - \varphi(X_i)\right) \\ &\quad + \left(\widehat{\varphi}(X_i) - \varphi(X_i)\right)^2 R_n^1(X_i) \\ &\quad + \left(\widehat{\varphi}(X_i) - \varphi(X_i)\right) \left(\widehat{\Phi}(X_i) - \Phi(X_i)\right) R_n^2(X_i) \end{aligned}$$

Dans cette expression on pose :

$$R_n^1(X_i) = -\frac{\alpha_n \widehat{\Phi}(X_i) + (1-\alpha_n)\Phi(X_i)}{\left(\alpha_n \widehat{\varphi}(X_i) + (1-\alpha_n)\varphi(X_i)\right)^3}$$

et

$$R_n^2(X_i) = \frac{-1}{2} \frac{1}{\left(\alpha_n \widehat{\varphi}(X_i) + (1-\alpha_n)\varphi(X_i)\right)^2}$$

en regroupant les termes, on obtient :

$$\widehat{f_n}(X_i) - f(X_i) = \frac{\widehat{\Phi}(X_i) - f(x_i) \cdot \widehat{\varphi}(X_i)}{\varphi(X_i)} + R_n(X_i)$$

où $\alpha_n \in [0,1]$, et où R_n est défini par :

$$\begin{aligned} R_n(X_i) &= \left(\widehat{\varphi}(X_i) - \varphi(X_i)\right)^2 R_n^1(X_i) \\ &\quad + \left(\widehat{\varphi}(X_i) - \varphi(X_i)\right) \left(\widehat{\Phi}(X_i) - \Phi(X_i)\right) R_n^2(X_i) \end{aligned}$$

F^+ s'écrit donc :

$$F^+ = \frac{\sqrt{n}}{n}\left(\sum_{i=1}^{n} Z_i\left(\frac{\widehat{\Phi}(X_i) - f(x_i)\cdot\widehat{\varphi}(X_i)}{\varphi(X_i)}\right)\right) + \frac{\sqrt{n}}{n}\sum_i Z_i \cdot R_n(X_i) \qquad (4.1)$$

Nous avons donc $F^+ = F_1^+ + F_2^+$
Montrons que $F_1^+ = \frac{\sqrt{n}}{n}\sum R_n(X_i)\cdot Z_i \xrightarrow{\mathcal{P}}$
$\longrightarrow 0$ pour cela on utilise la majoration suivante :

$$\left|\tfrac{\sqrt{n}}{n}\textstyle\sum_i Z_i\cdot\left(\widehat{\varphi}(X_i)-\varphi(X_i)\right)^2 R_n^1(X_i)\right|$$

$$\leq \sqrt{n}\left(Sup_{X_i}\ |\widehat{\varphi}(X_i)-\varphi(X_i)|\right)^2\cdot\tfrac{1}{n}\textstyle\sum|Z_i|\,|R_n^1(X_i)|$$

dans un voisinage compact de f , la fonction $|R_n^1(X_i)|$ est bornée par une fonction $\rho^1(X_i)$ indépendante de n, donc $\frac{1}{n}\sum|Z_i|\cdot\rho^1(X_i)$ est un $O_p(1)$.

De plus, $Sup_{X_i}\ |\widehat{\varphi}(X_i)-\varphi(X_i)|$ est d'ordre $Max\left(\frac{\sqrt{\log(n)}}{\sqrt{n\cdot h^p}}\ , h^d\right)$, donc sous l'hypothèse (4.9), $\sqrt{n}\left(Sup_{X_i}\ |\widehat{\varphi}(X_i)-\varphi(X_i)|\right)^2$ tend vers 0.

Le même traitement est appliqué au terme contenant R_n^2, ce qui nous donne le résultat annoncé.

Deuxième étape :

La normalité asymptotique provient du terme E^+ issu de notre décomposition initiale, auquel s'ajoute F_1^+ issu de (4.1), ce deux termes forment G^+ défini par :

$$G^+ = \frac{\sqrt{n}}{n}\cdot\sum_i\ Z_i\left(Y_i - f(X_i)\right) - \frac{\sqrt{n}}{n}\left(\sum_{i=1}^{n} Z_i\left(\frac{\widehat{\Phi}(X_i) - f(x_i)\cdot\widehat{\varphi}(X_i)}{\varphi(X_i)}\right)\right)$$

en utilisant l'expression non-paramétrique de $\widehat{\Phi}(X_i)$, nous obtenons :

$$\begin{aligned} G^+ \ &= \tfrac{\sqrt{n}}{n}\cdot\textstyle\sum_i\ Z_i\left(Y_i - f(X_i)\right) \\ &\quad - \tfrac{\sqrt{n}}{n}\cdot\textstyle\sum_{i=1}^n\tfrac{1}{n}\sum_{j=1}^n\tfrac{Z_i}{\varphi(X_i)}\left(Y_j - f(X_i)\right)\cdot\tfrac{1}{h^p}K\left(\tfrac{X_i-X_j}{h}\right) \end{aligned}$$

L'élimination du terme $i=j$ n'influence pas la limite de ce dernier terme, et on peut écrire :

$$\begin{aligned} G^+ \ &= \tfrac{\sqrt{n}}{n}\cdot\textstyle\sum_i\ Z_i\left(Y_i - f(X_i)\right) \\ &\quad - \tfrac{\sqrt{n}}{n\cdot(n-1)}\cdot\textstyle\sum_{i,j\neq i}\tfrac{Z_i}{\varphi(X_i)}\left(Y_j - f(X_i)\right)\cdot\tfrac{1}{h^p}K\left(\tfrac{X_i-X_j}{h}\right) \end{aligned} \qquad (4.2)$$

Un traitement par U-statistique (voir Serfling [79]) est appliqué au second terme, pour cela on pose :

$$U_n = \frac{1}{n\cdot(n-1)}\sum_{i,j\neq i}^{n}\lambda_2\left(S_i, S_j\right) \qquad (4.3)$$

avec $S_i = (X_i, Y_i, Z_i)$ et $\lambda_2(S_i, S_j)$ est la fonction symétrique :

$$\begin{aligned}\lambda_2(S_i, S_j) &= \tfrac{1}{2}\tfrac{Z_i}{\varphi(X_i)}(Y_j - f(X_i)) \cdot \tfrac{1}{h^p} K\left(\tfrac{X_i - X_j}{h}\right) \\ &+ \tfrac{1}{2}\tfrac{Z_j}{\varphi(X_j)}(Y_i - f(X_j)) \cdot \tfrac{1}{h^p} K\left(\tfrac{X_i - X_j}{h}\right)\end{aligned}$$

cette U-statistique est asymptotiquement équivalente à sa projection sur les observations $\widehat{U_n}$:

$$\widehat{U_n} = E[\lambda_2(S_i, S_j)] + \frac{1}{n}\sum_{i=1}^{n} \lambda_i(S_i) \tag{4.4}$$

où $\lambda_i(s)$ est :

$$\lambda_i(s) = E[\lambda_2(S_i, S_j) \mid S_i = s]$$

en décomposant le calcul de $\lambda_i(s)$

$$\lambda_i(s) = E[\lambda_2(S_i, S_j) \mid S_i = s] = E[E[\lambda_2(S_i, S_j) \mid X_i = x] \mid Y_i = y, Z_i = z]$$

on obtient :

$$\lambda_i(S_i) = m(X_i)(Y_i - f(X_i))$$

où $m(x) = E[Z \mid X = x]$

ceci nous amène à la conclusion car :

$$E[\lambda_n(S_i, S_j)] \longrightarrow 0$$

d'après (4.3) et (4.4)

$$\frac{\sqrt{n}}{n} \cdot \sum_{i,j \neq i}^{n} \lambda_n(\omega_i, \omega_j) \sim \frac{\sqrt{n}}{n} \cdot \sum_{i=1}^{n} m(X_i)(Y_i - f(X_i))$$

On obtient finalement en regroupant les termes de (4.2)

$$\sqrt{n} \cdot \left(\widehat{\gamma} - \widehat{\Gamma}(\widehat{f_n})\right) = Var(z)^{-1} \cdot \left[\frac{\sqrt{n}}{n}\sum_{i=1}^{n}(Z_i - m(X_i))(Y_i - f(X_i))\right] + O_p(1)$$

sous $\mathcal{H}_1$, le terme général de cette somme de termes *iid* est centré.

Par le même raisonnement que dans le cas PP, et sous l'hypothèse d'homoscédasticité 4.4 sa variance est $\sigma^2.E[Var(Z \mid X)]$, ce qui nous donne le résultat. □

Démonstration du théorème 4.7 (Statistique globale)

Etude de $\psi = \int (\delta_{f,g}(z))^2 \varpi(z) dz$

Par le choix de notre fonction de poids $\varpi(z) = \widehat{\varphi}^2(z) \cdot \nu_n(z)$ on a :

$$\psi = \frac{1}{n^2k^2} \int \left(\sum_i \left(Y_i - \widehat{f}(X_i) \right) K \left(\frac{Z_i - z}{k} \right) \right)^2 \nu_n(z) dz$$

que l'on décompose :

$$\begin{aligned} \psi &= \tfrac{1}{n^2k^2} \int \left(\textstyle\sum_i \left(Y_i - f(X_i) \right) K \left(\tfrac{Z_i - z}{k} \right) + \sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\tfrac{Z_j - z}{k} \right) \right)^2 \nu_n(z) dz \\ &= \tfrac{1}{n^2k^2} \int \left(Q_1(z) + Q_2(z) \right)^2 \nu_n(z) dz \end{aligned}$$

avec :

$$\begin{aligned} & Q_1(z) = \textstyle\sum_i \left(Y_i - f(X_i) \right) K \left(\tfrac{Z_i - z}{k} \right) \\ et & \\ & Q_2(z) = \textstyle\sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\tfrac{Z_j - z}{k} \right) \end{aligned}$$

D'où nous tirons **trois** termes, correspondant respectivement à l'intégrale de la somme des carrés de Q_1 et de Q_2, et du double produit :

$$F_1 = \tfrac{1}{n^2k^2} \int \left[\textstyle\sum_i \left(Y_i - f(X_i) \right) K \left(\tfrac{Z_i - z}{k} \right) \right]^2 \nu_n(z) dz$$

$$F_2 = \tfrac{1}{n^2k^2} \int \left[\textstyle\sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\tfrac{Z_j - z}{k} \right) \right]^2 \nu_n(z) dz$$

et

$$F_3 = \tfrac{1}{n^2k^2} \int \left[\textstyle\sum_i \left(Y_i - f(X_i) \right) K \left(\tfrac{Z_i - z}{k} \right) \right] \cdot \left[\textstyle\sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\tfrac{Z_j - z}{k} \right) \right] \nu_n(z) dz$$

Vue globale :

$$\psi = F_1 + F_2 + 2 \cdot F_3$$

- F_1 est explicitement étudié par Hall [45] et donne :

$$\begin{aligned} F_1 &= \tfrac{1}{nk} \sigma^2 \cdot \int K^2(u) \nu_n(u) du \\ &+ I_{n,2} \\ &+ O_p \left(\tfrac{1}{\sqrt{n^3 \cdot k^2}} \right) \end{aligned}$$

où

$$\left(n \cdot \sqrt{k} \right) \cdot I_{n,2} \; \mathcal{D}$$

$$\longrightarrow \mathcal{N} \left(0, \tfrac{1}{4} \cdot \alpha_1 \right)$$

- $n\sqrt{k}\cdot F_2 \to 0$ et l'on peut s'assurer de son comportement par un bon choix de la fenêtre "h" qu'il fait intervenir, car :

$$F_2 \leq O_p\left(Max\left(\frac{\log(n)}{n\cdot h}\ ,\ h^{2\cdot d}\right)\right)\cdot\int\widehat{\varphi}^2(z)\nu_n(z)dz$$

- $n\sqrt{k}\cdot F_3 \to 0$ assez lentement, on montre en effet que

$$n\sqrt{k}\cdot F_3 = O_p\left(k\right)$$

Etude de $\mathbf{F}_1$

Nous pouvons décomposer $\mathbf{F}_1$ en deux :

$$\begin{aligned} F_1 &= \tfrac{1}{n^2k^2}\int\left[\textstyle\sum_i\left(Y_i - f(X_i)\right)K\left(\frac{Z_i-z}{k}\right)\right]^2\nu_n(z)dz \\ &= \tfrac{1}{n^2k^2}\int\textstyle\sum_i\left(Y_i - f(X_i)\right)^2K^2\left(\frac{Z_i-z}{k}\right)\nu_n(z)dz \\ &+2\cdot\tfrac{1}{n^2k^2}\int\textstyle\sum_{i,j}\sum_{i<j}\left(Y_j - f(X_j)\right)K\left(\frac{Z_j-z}{k}\right)\left(Y_i - f(X_i)\right)K\left(\frac{Z_i-z}{k}\right)\nu_n(z)dz \\ &= I_{n,1} + 2\cdot I_{n,2} \end{aligned}$$

Ces deux termes $I_{n,1}$ et $I_{n,2}$ ont des comportements asymptotiques différents :

i)

$$I_{n,1} = \frac{1}{nk}\cdot\sigma^2\left(\int K^2(u)\nu_n(u)du\right) + O_p\left(\frac{1}{\sqrt{n^3\cdot k^2}}\right) \qquad (4.5)$$

tandis que

ii)

$$\left(n\cdot\sqrt{k}\right)\cdot I_{n,2}\ \mathcal{D}$$

$$\longrightarrow\mathcal{N}\left(0,\tfrac{1}{4}\cdot\alpha_1\right) \quad (4.6)$$

- **Etude de $I_{n,1}$:**

Nous noterons E' l'espérance conditionnelle aux X_i et aux Z_i :
Nous allons étudier tout d'abord :

$$(nk)^4\cdot E'\left[\left(I_{n,1} - E'\left[I_{n,1}\right]\right)^2\right] = \sum_i E'\left[\left(Y_i - f(X_i)\right)^2 - \sigma^2\right]^2\cdot\left[\int K^2\left(\frac{Z_i - z}{k}\right)\nu_n(z)dz\right]^2$$

On remarque que :

$$\int K^2\left(\frac{Z_i - z}{k}\right)\nu_n(z)dz \leq k\int K^2(u)\,\nu(u)du$$

sous l'hypothèse 4.12 et d'après la loi faible des grands nombres :

$$(nk)^4 \cdot E'\left[\left(I_{n,1} - E'[I_{n,1}]\right)^2\right] = O_p\left((nk)^2\right)\sum_i f_4(X_i) = O_p\left(nk^2\right)$$

Nous pouvons utiliser l'inégalité de Chebyshev, soit $(\lambda_n)_{n\in N}$ une suite divergente, on a :

$$\Pr\left\{|I_{n,1} - E'[I_{n,1}]| > \lambda_n \cdot \sqrt{\frac{nk^2}{(nk)^4}} \mid X_i, Z_i\right\} \overset{n\to\infty}{\longrightarrow} 0$$

et donc

$$I_{n,1} = E'[I_{n,1}] + O_p\left(\sqrt{\frac{nk^2}{(nk)^4}}\right)$$

Ce qui montre i)

- **Etude de $I_{n,2}$**

Le second terme $I_{n,2}$, est celui qui nous donne la normalité asymptotique.

$$(nk)^2 I_{n,2} = \frac{1}{n^2k^2}\sum_{i,j}\sum_{i<j}(Y_j - f(X_j))(Y_i - f(X_i))\int K\left(\frac{Z_j - z}{k}\right)K\left(\frac{Z_i - z}{k}\right)\nu_n(z)dz$$

Nous pouvons réécrire cette expression :

$$(nk)^2 I_{n,2} = \sum_{i,j}\sum_{i<j}(Y_j - f(X_j))(Y_i - f(X_i))W_{n,i,j}$$

ou encore

$$(nk)^2 I_{n,2} = \sum_{i=2}^{n} Y_{n,i} \tag{4.7}$$

avec

$$Y_{n,i} = (Y_i - f(X_i))\sum_{j=1}^{i-1}(Y_j - f(X_j))W_{n,i,j} \quad 2 \leq i \leq n$$

Soit $\mathcal{F}_{n,i}$ la tribu engendrée par $S_1, \cdots, S_n$ (où $S_i = (X_i, Y_i, Z_i)$) alors

$$E[Y_{n,i} \mid \mathcal{F}_{n,i}] = 0 \quad p.s. \quad \forall i$$

et la séquence

$$\left\{\left(M_{n,i} = \sum_{j=2}^{i} Y_{n,j}\ , \mathcal{F}_{n,i}\right) \quad , 2 \leq i \leq n\right\}$$

est une martingale. La variance conditionnelle de $M_{n,n}$ est :

$$V_n = \sum_{i=2}^{n} E\left[Y_{n,i}^2 \mid \mathcal{F}_{n,i-1}\right] = \sum_{i=2}^{n} \sigma^2 \left[\sum_{j=1}^{i-1} \left(Y_j - f(X_j)\right) W_{n,i,j}\right]^2$$

On la décompose en deux termes (la somme des carrés et le double produit) :

$$\begin{aligned} V_n &= n.\sigma^2 \textstyle\sum_{j=1}^{i-1} \left(Y_j - f(X_j)\right)^2 W_{n,i,j}^2 \\ &+2n\sigma^2 \textstyle\sum_{1\leq j\leq l\leq i-1} \left(Y_j - f(X_j)\right)\left(Y_l - f(X_l)\right) W_{n,i,j} W_{n,i,l} \\ &= V_{n,1} + V_{n,2} \end{aligned}$$

Hall ([45] lemme 1 et 2) sur les base d'un théorème central limite dû à Brown [14], nous indique que :

$$\frac{1}{n \cdot k^{3/2}} \cdot M_{n,n} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}\left(0, \tfrac{1}{4}\alpha_1\right)$$

où α_1 est défini par :

$$\frac{1}{n^2 \cdot k^3} \cdot V_{n,1} \longrightarrow \frac{1}{4}\alpha_1$$

et

$$\alpha_1 = 2 \cdot \sigma^4 \left[\int \varphi^2(x)\nu(x)dx\right] \cdot \left[\int \left[\int K(u)K(u+v)du\right]^2 dv\right]$$

En reportant ce résultat dans (4.7), nous obtenons que :

$$\frac{1}{n \cdot k^{3/2}} \cdot M_{n,n} = \left(n \cdot \sqrt{k}\right) \cdot I_{n,2} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}\left(0, \tfrac{1}{4} \cdot \alpha_1\right)$$

ce qui montre (1.4.1) et nous donne ii).

Au total :

$$\begin{aligned} F_1 &= \tfrac{1}{nk}\, \sigma^2 \cdot (\cdot K^2(u)\nu_n(u)du) \\ &+2 \cdot I_{n,2} \\ &+O_p\left(\tfrac{1}{\sqrt{n^3 \cdot k^2}}\right) \end{aligned}$$

$\square_1$

Etude de $\mathbf{F}_2$

Nous cherchons à faire disparaître asymptotiquement le terme $n \cdot \sqrt{k} \cdot F_2$, où F_2 est défini par :

$$F_2 = \frac{1}{n^2 k^2} \int \left[\sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\frac{Z_j - z}{k} \right) \right]^2 \nu_n(z) dz$$

L'étude de ce terme est simplifié grandement par notre hypothèse 4.14, car

$$F_2 \leq \left(\sup_{X_j} \left| f(X_j) - \widehat{f}(X_j) \right| \right)^2 \int \left[\frac{1}{nk} \cdot \sum_j K \left(\frac{Z_j - z}{k} \right) \right]^2 \nu_n(z) dz$$

et

- $\left(\sup_{X_j} \left| f(X_j) - \widehat{f}(X_j) \right| \right)^2 \longrightarrow 0$
- $\int \left[\frac{1}{nk} \cdot \sum_j K \left(\frac{Z_j - z}{k} \right) \right]^2 \nu_n(z) dz \longrightarrow \int \varphi^2(z) \cdot \nu_n(z) dz$

On a donc :

$$\left(\sup_{X_j} \left| f(X_j) - \widehat{f}(X_j) \right| \right)^2 = O_p \left(Max \left(\frac{\log(n)}{n \cdot h} \ , \ h^{2 \cdot d} \right) \right)$$

D'où nous tirons :

$$\left(n \cdot \sqrt{k} \right) \cdot F_2 \leq O_p \left(Max \left(\frac{n \cdot \sqrt{k} \cdot \log(n)}{n \cdot h} \ , n \cdot \sqrt{k} \ h^4 \right) \right) \cdot \int \widehat{\varphi}^2(z) \nu_n(z) dz$$

Sous les hypothèses 4.13 et 4.14, on vérifie que :

$$\frac{n \cdot \sqrt{k} \cdot \log(n)}{n \cdot h} \overset{n \to \infty}{\longrightarrow} 0$$

et

$$n \cdot \sqrt{k} \ h^4 \leq n \cdot h^5 \overset{n \to \infty}{\longrightarrow} 0$$

Le terme $\left(n \cdot \sqrt{k} \right) \cdot F_2$ converge donc bien vers 0 en probabilité. $\square_2$

Etude de $\mathbf{F}_3$

$$F_3 = \frac{1}{n^2 k^2} \int \left[\sum_i \left(Y_i - f(X_i) \right) K \left(\frac{Z_i - z}{k} \right) \right] \cdot \left[\sum_j \left(f(X_j) - \widehat{f}(X_j) \right) K \left(\frac{Z_j - z}{k} \right) \right] \nu_n(z) dz$$

$$= \frac{1}{n^2 k} \sum_i \sum_j \left(Y_i - f(X_i) \right) \cdot \left(f(X_j) - \widehat{f}(X_j) \right) \cdot \int \frac{1}{k} K \left(\frac{Z_i - z}{k} \right) K \left(\frac{Z_j - z}{k} \right) \nu_n(z) dz$$

Le terme impliquant le produit des noyaux peut être vu comme

$$\int \frac{1}{k} \cdot \Upsilon(z) K\left(\frac{Z_j - z}{k}\right) dz$$

et est équivalent asymptotiquement à

$$K\left(\frac{Z_j - Z_i}{k}\right) \nu(Z_i)$$

Nous allons donc nous concentrer sur l'étude de

$$\widetilde{F_3} = \frac{1}{n^2 k} \sum_i \sum_j (Y_i - f(X_i)) \cdot \left(f(X_j) - \widehat{f}(X_j)\right) \cdot K\left(\frac{Z_j - Z_i}{k}\right) \nu(Z_i)$$

Posons $\forall i = 1, \cdots, n \quad \varepsilon_i = Y_i - f(X_i)$. Afin de montrer que $n \cdot \sqrt{k}$ $\widetilde{F_3} \to 0$ en probabilité, nous allons étudier les deux premiers moments de $\widetilde{F_3}$. Clairement

$$E\left[\widetilde{F_3}\right] = 0$$

Etudions $\widetilde{F_3}^2$

$$\widetilde{F_3}^2 = \frac{1}{n^4 k^2} \left(\sum_i \sum_j \varepsilon_i \cdot \left(f(X_j) - \widehat{f}(X_j)\right) \cdot K\left(\frac{Z_j - Z_i}{k}\right) \nu(Z_i) \right)^2$$

avant de décomposer cette somme nous pouvons introduire l'expression de $f(X_j) - \widehat{f}(X_j)$ et obtenir la somme triple (au carré) :

$$\widetilde{F_3}^2 = \frac{1}{n^4 k^2} \left(\sum_i \sum_j \frac{1}{nh} \sum_l \varepsilon_i \cdot \varepsilon_l \cdot K\left(\frac{X_l - Xj}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right) \frac{\nu(Z_i)}{\widehat{\varphi}(X_j)} \right)^2$$

qui s'écrit en développant :

$$\begin{aligned} \widetilde{F_3}^2 \quad &= \tfrac{1}{n^4 k^2} \tfrac{1}{n^2 h^2} \sum_{i,j,l,i',j'l'} \varepsilon_i \varepsilon_{i'} \varepsilon_l \varepsilon_{l'} \cdot K\left(\tfrac{X_l - Xj}{h}\right) K\left(\tfrac{X_{l'} - Xj'}{h}\right) \\ &\cdot K\left(\tfrac{Z_j - Z_i}{k}\right) K\left(\tfrac{Z_{j'} - Z_{i'}}{k}\right) \cdot \tfrac{\nu(Z_i)}{\widehat{\varphi}(X_j)} \cdot \tfrac{\nu(Z_{i'})}{\widehat{\varphi}(X_{j'})} \end{aligned}$$

L'intérêt de cette décomposition est qu'il y est plus facile de voir que l'espérance des termes constituant cette somme est nulle, sauf dans un nombre réduit de cas. En d'autres termes :

$$E\left[\varepsilon_i \varepsilon_{i'} \varepsilon_l \varepsilon_{l'} \cdot K\left(\frac{X_l - Xj}{h}\right) K\left(\frac{X_{l'} - Xj'}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right) K\left(\frac{Z_{j'} - Z_{i'}}{k}\right) \cdot \frac{\nu(Z_i)}{\widehat{\varphi}(X_j)} \frac{\nu(Z_{i'})}{\widehat{\varphi}(X_{j'})}\right] = 0$$

Sauf

- pour le cas où $i = i' = l = l'$, nous aurons alors la somme sur les trois indices i, j et j' de cette espérance, (cas A)

- pour les 3 cas où l'on a des "couples" C_1 : $(i = i' \ \ et \ \ l = l')$; C_2 : $(i = l \ \ et \ \ i' = l')$ et C_3 : $(i = l' \ \ et \ \ i' = l)$, où nous aurons une somme sur 4 indices (j, j', i et l dans le premier cas)

Nous avons donc

$$\begin{aligned} E\left[\widetilde{F_3}^2\right] &= \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{i,j,j'} E\left[A\right] \\ &+\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{j,j',i,l}\ E\left[C_1\right] \\ &+\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{j,j',i,i'}\ E\left[C_2\right] \\ &\tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{j,j',i,l}\ E\left[C_3\right] \end{aligned}$$

Etudions tout d'abord

$$\begin{aligned} \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{i,j,j'} E\left[A\right] &= \tfrac{1}{n^4}\left(\textstyle\sum_i f_4(X_i)\right. \\ &\times \textstyle\sum_{j,j'} \tfrac{1}{nh^2} K\left(\tfrac{X_i - Xj}{h}\right) K\left(\tfrac{X_i - Xj'}{h}\right) \\ &\left.\cdot \tfrac{1}{nk^2} K\left(\tfrac{Z_j - Z_i}{k}\right) K\left(\tfrac{Z_{j'} - Z_i}{k}\right) \cdot \tfrac{\nu(Z_i)}{\widehat{\varphi}(X_j)} \tfrac{\nu(Z_i)}{\widehat{\varphi}(X_{j'})}\right) \end{aligned}$$

en réorganisant les termes on a :

$$\begin{aligned} \tfrac{1}{n^6k^2}\tfrac{1}{h^2}\ \textstyle\sum_{i,j,j'} E\left[A\right] &= \tfrac{1}{n^4}\left(\textstyle\sum_i f_4(X_i)\right. \\ &\times \textstyle\sum_{j} \tfrac{1}{nh^2} K\left(\tfrac{X_i - Xj}{h}\right) K\left(\tfrac{Z_j - Z_i}{k}\right) / \widehat{\varphi}(X_j) \\ &\left.\cdot \textstyle\sum_{j'} \tfrac{1}{nk^2} K\left(\tfrac{X_i - Xj'}{h}\right) K\left(\tfrac{Z_{j'} - Z_i}{k}\right) / \widehat{\varphi}(X_{j'}) \cdot \nu^2(Z_i)\right) \end{aligned}$$

le terme

$$\sum_j \frac{1}{nh^2} K\left(\frac{X_i - Xj}{h}\right) K\left(\frac{Z_j - Z_i}{k}\right) / \widehat{\varphi}(X_j) \longrightarrow \frac{\varphi(X_i, Z_i)}{\varphi(X_i)} = \varphi(Z_i \mid X_i)$$

et est supposé borné, d'où d'après 4.12 et la loi des grands nombres.

$$\frac{1}{n^6k^2}\frac{1}{h^2}\ \sum_{i,j,j'} E\left[A\right] = O_p(1) \cdot \frac{1}{n^4} \sum_i f_4(X_i)\nu^2(Z_i) = O_p\left(\frac{1}{n^3}\right)$$

nous étudions ensuite le terme $\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{j,j',i,l} E[C_1]$,

$$\begin{aligned}\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{i,j,j'} E[C_1] &= \frac{1}{n^4}\Big(\sum_{i,l}\sigma^2\sigma^2 \\ &\quad \times \sum_{j,j'}\frac{1}{nh^2}K\left(\frac{X_i-Xj}{h}\right)K\left(\frac{X_i-Xj'}{h}\right) \\ &\quad \cdot \frac{1}{nk^2}K\left(\frac{Z_j-Z_i}{k}\right)K\left(\frac{Z_{j'}-Z_i}{k}\right)\cdot\frac{\nu(Z_i)}{\widehat{\varphi}(X_j)}\frac{\nu(Z_i)}{\widehat{\varphi}(X_{j'})}\Big)\end{aligned}$$

de la même manière que précédemment on peut regrouper les sommes, pour obtenir une expression semblable,

$$\begin{aligned}\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{i,j,j'} E[C_1] &= \frac{1}{n^4}\Big(\sum_{i,l}\sigma^4 \\ &\quad \times \sum_{j}\frac{1}{nh^2}K\left(\frac{X_i-Xj}{h}\right)K\left(\frac{Z_j-Z_i}{k}\right)/\widehat{\varphi}(X_j) \\ &\quad \cdot \sum_{j'}\frac{1}{nk^2}K\left(\frac{X_i-Xj'}{h}\right)K\left(\frac{Z_{j'}-Z_i}{k}\right)/\widehat{\varphi}(X_{j'})\cdot\nu^2(Z_i)\Big)\end{aligned}$$

nous obtenons cette fois

$$\frac{1}{n^6k^2}\frac{1}{h^2}\sum_{i,j,j'} E[C_1] = O_p(1)\cdot\frac{1}{n^4}\sigma^4\sum_{i,l}\nu^2(Z_i) = O_p\left(\frac{1}{n^2}\right)$$

Au total

$$E\left[\widetilde{F_3}^2\right] = O_p\left(\frac{1}{n^3}+\frac{1}{n^2}\right) = O_p\left(\frac{1}{n^2}\right)$$

d'où

$$E\left[\left(n\sqrt{k}.F_3\right)^2\right] = O_p(k)$$

Ce qui donne le résultat annoncé pour F_3 $\square_3$

Nous obtenons le résultat final par addition des trois termes. $\square$
[12pt,qqa4lrep]report

Chapter 5

Simulations

Nous présentons ici, dans le cas univarié, les premiers résultats de simulation des théorèmes asymptotiques formulés section 4.3.

Le protocole de simulation vise à engendrer les vecteurs ou matrices nous permettant de visualiser le comportement asymptotique de nos statistiques. Ces statistiques sont toutes basées sur la différence entre un estimateur du paramètre (scalaire ou fonctionnel) identifiant $\mathcal{M}_2$ et un estimateur de la pseudo-vraie valeur. Le calcul répété de ces statistiques nous donne des vecteurs correspondant à N tirages différents de la statistique considérée. Nous construisons la répartition empirique de cette statistique sur les bases de ce vecteur. Nous présenterons ces résultats sous forme d'histogrammes.

Comme nous l'avons vu, ces statistiques peuvent revêtir une forme paramétrique ou fonctionnelle suivant la nature paramétrique ou fonctionnelle du modèle $\mathcal{M}_2$. Nous construirons donc également des histogrammes en plusieurs points pour la visualisation des statistiques fonctionnelles.

5.1 Protocole de simulation des données

Nous avons conduit quatre procédures de génération des données permettant de calculer nos statistiques suivant les différents aspects que peuvent revêtir les observations. Ces quatre procédures seront désignées dans tout ce chapitre par les lettres A, B, C, et D.

La procédure A :

Pour tirer les vecteurs (X_i, Y_i, Z_i) nous utiliserons la procédure :

- On tire les vecteurs $(X_i, Z_i) \sim \mathcal{IN}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

- Le vecteur Y_i est lui calculé à partir de l'expression :

$$y = \eta \cdot x + \cdot u$$

  - où $u \sim \mathcal{N}(0, s_y)$; Y_i est alors issu d'un modèle linéaire standard.

Cette procédure de simulation des données est réglée par le choix de trois paramètres, le paramètre ρ de covariance de (X_i, Z_i), le coefficient linéaire η et la variance du résidu s_y.

- Dans la pratique nous avons choisi $= \rho = 0.5$, $\eta = 2$ et $s_y = 0.5$.

La procédure B :

Pour tirer les vecteurs (X_i, Y_i, Z_i) nous utiliserons la procédure :

- On tire d'abord le vecteur $X_i \sim \mathcal{IN}(0, 1)$
- puis le vecteur Z_i relié au vecteur X_i par la relation :

$$z = \frac{4}{1 + e^{-a \cdot x}} - \frac{1}{2} + u$$

  - où $u \sim \mathcal{N}(0, s_z)$

- Le vecteur Y_i s'exprime de manière non-linéaire en fonction de x par :

$$y = \lambda \cdot x + \frac{1}{1 + e^{-b \cdot x}} - \frac{1}{2} + v$$

  - où $v \sim \mathcal{N}(0, s_y)$.

Cette procédure de simulation des données est conditionnée par le choix des paramètres a, b, λ et les deux paramètres "*de bruit*" s_y et s_z.

Dans la pratique nous avons été amenés à choisir ces coefficients prenant les valeurs $a = 3$, $b = 2$, $\lambda = 1$ les deux paramètres s_y et s_z étant fixés à 0.5

La procédure C :

Nous reprenons ici certains des points concernant le tirage du couple (X_i, Y_i, Z_i) proposés ci-dessus, en mixant les procédures A et B

- On tire d'abord le vecteur $X_i \sim \mathcal{IN}(0, 1)$

- puis le vecteur Z_i relié au vecteur X_i par la relation :

$$z = \frac{4}{1 + e^{-a \cdot x}} - \frac{1}{2} + u$$

  - où $u \sim \mathcal{N}(0, s_z)$

- Nous retrouvons ici le premier point de la procédure B, toutefois le vecteur Y_i s'exprime conformément à la procédure A, par :

$$y = \eta \cdot x + \cdot u$$

  - où $u \sim \mathcal{N}(0, s_y)$.

Les paramètres déterminant cette procédure sont a, η et les deux paramètres "*de bruit*" s_y et s_z.

Dans la pratique $a = 3, \eta = 2$ les deux paramètres s_y et s_z sont fixés à 0.5.

La procédure D :

Il nous reste une possibilité de croisement des procédures A et B pour tirer les vecteurs (X_i, Y_i, Z_i) :

- On tire les vecteurs $(X_i, Z_i) \sim \mathcal{IN}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$

- Le vecteur Y_i est lui calculé à partir de l'expression :

$$y = \lambda \cdot x + \frac{1}{1 + e^{-b \cdot x}} - \frac{1}{2} + v$$

  - où $v \sim \mathcal{N}(0, s_y)$.

Dans cette procédure de simulation les paramètres ρ, λ, b et les deux paramètres "*de bruit*" s_y et s_z seront identiques aux choix effectués précédemment, à savoir :
$\rho = 0.5,\ \lambda = 1,\ b = 2,$ et $s_z = s_y = 0.5$.

5.2 Algorithme

L'algorithme de simulation que nous avons utilisé est relativement classique, il comporte trois grandes phases :

- Simulation des données

- Calcul des différents estimateurs
- Calcul des statistiques

Un listing des programmes Gauss est donné en annexe, sa lecture n'étant pas des plus agréables, l'algorithme du programme général est rapporté ici[1] :

Pour t=1 à n :

Simulation des données

Calcul des estimateurs paramétriques $\widehat{\beta}$ et $\widehat{\gamma}$.

Calcul des fenêtres optimales h_{opt} et k_{opt} pour les estimateurs non-paramétriques.

Pour b=1 à B

- Calcul de l'estimateur de leave-one-out
- Calcul du critère de validation croisée $CV(h)$ et $CV(k)$
- Stockage de $CV(h)$ et $CV(k)$
- b=b+1 (On change de fenêtre)

Choix du critère minimal déterminant h_{opt}

Même procédure pour k_{opt}

Calcul des estimateurs non-paramétriques $\widehat{f_n}$ et $\widehat{g_n}$.

Calcul des pseudo-vraies valeurs

Calcul des variances asymptotiques

Calcul des statistiques

- t=t+1 (On tire un nouvel échantillon)

Stockage des données et graphique

[1] Il y a en fait quatre programmes correspondant aux quatre procédures de simulation des données.

5.3 Statistique $\delta_{\beta,\gamma}$

La statistique paramétrique $\sqrt{n}\cdot\delta_{\beta,\gamma} = \sqrt{n}\cdot\left(\widehat{\gamma} - \widehat{\Gamma_L}(\widehat{\beta})\right)$ est issue du théorème 4.3 et est distribuée asymptotiquement selon une loi normale centrée de variance $\sigma^2 \cdot V(\delta)$. Nous présentons dans les figures **??**, 5.1, 5.2 et 5.3 les histogrammes représentant la densité de

$$\sqrt{n} \cdot \frac{\delta_{\beta,\gamma}}{\widehat{\sigma} \cdot \sqrt{\widehat{V(\delta)}}}$$

effectués sur la base de nos simulations.

Les simulations présentées portent sur des échantillons de taille 300 ; étant donné la simplicité des expressions intervenant dans le calcul de cette statistique, le nombre N de réplications est ici de 1000.

ftbpFU1pt0pt0pt Densité de la statistique PP, Données A.appFigure

Figure 5.1: Densit de la statistique PP, Donnes B.

Figure 5.2: Densit de la statistique PP, Donnes C.

Figure 5.3: Densit de la statistique PP, Donnes D.

5.4 Statistique $\delta_{f,g}$

La statistique fonctionnelle $\delta_{f,g}(z)$ est définie à partir des estimateurs des régressions f et g et prend la forme

$$\sqrt{nk} \cdot \delta_{f,g} = \sqrt{nk} \cdot \left(\widehat{g_n}(z) - \widehat{G_f}(z) \right)$$

Elle est issue du théorème 4.4 et est distribuée asymptotiquement selon une loi normale centrée. Nous avons simulé cette statistique sur des échantillons de taille 300 et avons effectué 100 réplications.

Nous présentons plusieurs figures pour chaque protocole de simulation. Nous avons effectué 12 histogrammes pour 12 valeurs de z afin de visualiser la statistique en des points différents. Les fenêtres h et k ont été choisies par validation croisée, tandis que m restait égale à k. Cette dernière restriction est due à des problèmes de temps de calcul. Nous présentons ces simulations sur les figures 5.4 et suivantes pour la procédure A, ces simulations ont également été réalisées pour les procédures B, C et D, et donnent des résultats semblables quoique moins réguliers suivant les valeurs de z.

Nous avions également effectué des simulation du critère global empirique Ψ, défini par :

$$\Psi = \sum_{i=1}^{n} \left(\delta_{f,g}(Z_i) \right)^2$$

Ces représentations sont présentés dans les figures 5.7, et 5.8 pour les protocoles A et B, et pour des fenêtres optimales, et en 5.9 et 5.10 pour des fenêtres calculées par la formule

$$h_0 = M_t \cdot n^{-\frac{1}{5}} \quad et \quad k_0 = N_t \cdot n^{-\frac{1}{5}}$$

où M_t et N_t sont les écarts types des X_i et Z_i pour l'échantillon t.

Nous voulions montrer par ce calcul l'importance du décalage entre une statistique avec des fenêtres optimales et une statistique avec des fenêtres théoriques. Ce décalage est particulièrement sensible sur les données B (figures 5.8 et 5.10).

Figure 5.4: Statistique $\delta_{f,g}$

Figure 5.5: Statistique $\delta_{f,g}$

Figure 5.6: Statistique $\delta_{f,g}$

Figure 5.7: Statistique $\Psi(\delta_{f,g})$

Figure 5.8: Statistique $\Psi(\delta_{f,g})$

Figure 5.9: Statistique $\Psi(\delta_{f,g})$

Figure 5.10: Statistique $\Psi(\delta_{f,g})$

5.5 Statistique $\delta_{f,\gamma}$

La statistique paramétrique

$$\sqrt{n} \cdot \delta_{f,\gamma} = \sqrt{n} \cdot \left(\widehat{\gamma} - \widehat{\Gamma_f}(\widehat{\beta}) \right)$$

est issue du théorème 4.6 et est distribuée asymptotiquement selon une loi normale centrée de variance $\sigma^2 \cdot V(\delta)$.

Nous présentons dans les figures **??**, et suivantes les histogrammes représentant la densité de

$$\sqrt{n} \cdot \frac{\delta_{f,\gamma}}{\widehat{\sigma} \cdot \sqrt{\widehat{V(\delta)}}}$$

effectués sur la base de nos simulations.

Ces représentations ne sont pas de bonne qualité pour aucun des trois protocoles effectués. Nous notons une mauvaise estimation du biais, qui se visualise par un décalage vers la droite de nos histogrammes.

L'unique fenêtre h intervenant dans ces calculs ayant été estimée par validation croisée, deux éléments (au moins) peuvent être la cause de ces résultats décevants

- La taille de l'échantillon, qui est ici de 300 et qui est faible pour l'estimation de la régression de f
- Le nombre de réplication qui n'est que de 100, alors qu'il est de 1000 pour la statistique PP

Ces deux éléments sont donc à varier dans une future étude afin de préciser la forme de la distribution obtenue.

Figure 5.11: Statistique $\delta_{f,\gamma}$

Figure 5.12: Statistique $\delta_{f,\gamma}$

Figure 5.13: Statistique $\delta_{f,\gamma}$

5.6 Statistique $\delta_{\beta,f}$

Cette dernière statistique est fonctionnelle ; $\delta_{\beta,f}(z)$ est définie à partir des estimateurs $\widehat{\beta}$ et $\widehat{g}$ et prend la forme

$$\sqrt{nk} \cdot \delta_{\beta,f} = \sqrt{nk} \cdot \left(\widehat{g_n}(z) - \widehat{G_\beta}(z) \right)$$

Elle est issue du théorème 4.5 et est distribuée asymptotiquement selon une loi normale centrée. Nous avons simulé cette statistique sur des échantillons de taille 300 et avons effectué 100 réplications.

De même que précédemment ,nous présentons plusieurs figures pour chaque protocole de simulation. Nous avons effectué 12 histogrammes pour 12 valeurs aléatoires de z afin de visualiser la statistique en des points différents. La fenêtre k a été choisie par validation croisée, tandis que m restait égale à k.

Nous présentons ces simulations sur les figures 5.14 et suivantes pour la procédure A, les procédures C et D donnent des résultats semblables, la procédure B donnant elle des résultats plus mauvais.

Nous notons que la distribution empirique est assez bien centrée, mais qu'elle présente des irrégularités suivant les composantes de z. La fenêtre k a été estimée par validation croisée tandis que la pseudo-vraie fenêtre m restait neutre puisque fixée à $m = k$.

Figure 5.14: Statistique $\delta_{\beta,f}$

Figure 5.15: Statistique $\delta_{\beta,f}$

Figure 5.16: Statistique $\delta_{\beta,f}$

5.7 Conclusion

Les simulations présentées dans ce chapitre ne sont qu'un aperçu de futures simulations complètes que nous réaliserons en faisant varier l'ensemble des paramètres : taille de l'échantillon, nombre de réplications, nombre de protocoles utilisés, paramètres de ces protocoles, etc...

Même si ces simulations n'ont pas la qualité nécessaire pour pouvoir prétendre valider nos résultats, elles furent un élément utile à la compréhension des phénomènes intervenant dans nos statistiques. C'est grâce à ces représentations que nous avons mis en évidence les comportement dûs aux différentes fenêtres.

Les simulations portant sur la statistique globale étudié dans le théorème 4.7, sont encore à faire. Il s'agira de construire la distribution de cette statistique à partir d'un échantillon (X_i^*, Y_i^*, Z_i^*) créé par bootstrap, c'est à dire à partir d'un échantillon original de taille fixé $(X_i, Y_i, Z_i)_{i=1,\ldots,n}$. Toutefois Härdle et Mammen [49], nous mettent en garde sur la façon de Bootstraper, c'est à dire de "tirer" les éléments parmi l'échantillon original. Le "*Bootstrap naïf*", qui consiste à des retirages simples, avec remise, d'un triplet parmi les n originaux, ne s'avèrant pas satisfaisant dans leur contexte, une autre technique le "*wild bootstrap*" est alors proposée. Ces techniques constituent une alternative à la classique démarche asymptotique et seront les outils de validation de nos résultats, ils seront l'objet de travaux futurs.

[12pt,qqa4lrep]report

Bibliography

[1] Aït-Sahalia, Y. (92) : "The Delta and Bootstrap methods for nonlinear functionals of nonparametric kernel estimators based on dependent multivariate data", *Mimeo*, Department of Economics, MIT.

[2] Amemiya T. (80) : "Selection of regressors", *International Economic Review*, 21(2), pp. 331-354.

[3] Amemiya T. (85) : "Advanced Econometrics", *Basil Blackwell*, Oxford.

[4] Atkinson A.C. (70) : "A Method for Discriminating Between Models", *Journal of the Royal Statistical Society*, series B, no 32, pp. 323-344.

[5] Bickel P. J. and M. Rosenblatt (73) : "On Some Global Measures of the Deviations of Density Function Estimates", *Annals of Statistics,* Vol. 1, no 6, pp. 1071-1095.

[6] Bierens H.J. (83) : "Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions", *Journal of the American Statistical Association,* Vol.78, No. 383.

[7] Bierens H.J. (87) : " Kernel estimators of regression functions " in *Advances in econometrics,* Fifth World Congres, Vol.1, T.F. Bewley.

[8] Bierens H.J. (90) : "A Consistent Conditional Moment Test of Functional Form ", *Econometrica*, vol.58, no 6.

[9] Bierens H.J. et W. Ploberger (93) : "Asymptotic Optimality and Size of the Integrated Consistent Conditional moment Test of Functional Form", *Working paper*, Southern Methodist University, Dallas.

[10] Billingsley P. (86) : "Probability and measure", *Wiley & Sons.*

[11] Bochner S. (55) : "Harmonic Analysis and the Theory of Probability", *University of California Press.*

[12] Bosq D. et J.F. Lecoutre (87) : " Théorie de l'estimation fonctionnelle" *Economica.*

[13] Bouoiyour J. (93) : " Tests Bayesiens d'enveloppement et fonction de coût", *Thèse de Doctorat,* Université des Sciences Sociales de Toulouse.

[14] Brown B. M. (71) : "Martingale Central Limit Theorems", *Annals of Mathematical Statistics,* Vol 42, no 1, pp. 59-66.

[15] Carrasco M. (94) : " The asymptotic Distribution of the Wald Statistic in Misspecified Structural Change, Threshold or Markov Switching Models", *Mimeo*, Gremaq, Université des Sciences Sociales de Toulouse.

[16] Cristóbal Cristóbal J. A., P. F. Roca and W.G. Manteiga (87) : " A Class of Linear Regression Parameter Estimators Constructed by Nonparametric Estimation" *Annals of Statistics,* Vol 15, no 2, pp. 603-609.

[17] Collomb G. (76) : " Estimation non-paramétrique de la régression par la méthode du noyau", *Thèse*, Université Paul Sabatier, Toulouse.

[18] Collomb G. (77) : " Estimation Non-paramétrique de la Régression par la méthode du Noyau : propriété de convergence asymptotiquement normale indépendante", *Annales Scientifiques de l'Université de Clermont*, Vol. 15, pp.24-26.

[19] Collomb G. (81) : " Estimation Non-paramétrique de la Régression : Revue Bibliographique", *International Statistical Review*, 49, pp. 75-93.

[20] Collomb G. (85) : " Nonparametric Regression : An Up-To-date Bibliography", Statistics, Vol. 16, no 2, pp. 309-324.

[21] Cox D.R.(61) : " Tests of Separate Families of Hypotheses" in *Proceeding of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1,University of California Press, Berkeley, pp. 105-123.

[22] Cox D.R.(62) : "Further results on tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society*, Series B, no. 24, pp. 406-424.

[23] Davidson R. et J.G. Mac Kinnon (81) : "Several test for model specification in the presence of alternative hypotheses", Econometrica, vol. 49,pp. 781-793.

[24] Davidson R. et J.G. Mac Kinnon (93) : " Estimation and inference in econometrics ", *Oxford University Press.*

[25] De Jong P. (87) : " A Central Limit Theorem for Generalized Quadratic Forms", *Probability Theory and Related fields,* no 75, pp.261-277.

[26] Delecroix M. (83) : "Histogrammes et Estimation de la Densité", *Que sais-je ?* No. 2055, Presses Universitaires de France.

[27] Dhaene G. (93) : " Encompassing : Formulation, properties and testing " Ph. D Dissertation Université Catholique de Louvain, Louvain-la-Neuve.

[28] Dhaene G. (93) : " On the Encompassing Relation ", *Mimeo,* Université Catholique de Louvain, Louvain-la-Neuve.

[29] Engle R.F. , D.F.Hendry and J.F. Richard (83) : "Exogeneity", Econometrica, No. 55, pp 277-304.

[30] Eubank R. L. (88) : "Smoothing Splines and Nonparametric Regression", *Marcel Dekker*, New York.

[31] Florens J.P., D.F. Hendry and J. F. Richard (94) : " Encompassing and Specificity " *Mimeo*, Gremaq, Université des Sciences Sociales de Toulouse.

[32] Florens J.P. et S. Larribeau (91) : " Best approximated regression based on the behavior of the explanatory variables " *Mimeo*, Gremaq, Université des Sciences Sociales de Toulouse.

[33] Florens J.P., S. Larribeau et M. Mouchart (92) : " Bayesian encompassing tests of a unit root hypothese." *Mimeo*, Gremaq, No 92.27.

[34] Gasquet C. et P Witomski (89) : "Analyse de Fourier et Applications : Filtrage, Calcul Numérique, Ondelettes." Masson.

[35] Gasser T. and H. G. Müller (84) : "Estimating Regression Function and their Derivatives by the Kernel Method", *Scandinavian Journal of Statistics*, Vol. 11, pp. 171-185.

[36] Gasser T., A. Kneip and W. Köhler (91) : "A Flexible and Fast Method for Automatic Smoothing", *Journal of the Royal Statistical Society*, Series B, Vol.86, no. 415, pp. 643-652.

[37] Gourieroux C. et A. Monfort (89) : "Statistique et modèles économétriques", Vol. 1 et 2, *Economica*, Paris.

[38] Gourieroux C. et A. Monfort (91) : "Testing Non Nested Hypotheses ", *Mimeo* No 9207, Crest-Cepremap-Insee.

[39] Gourieroux C. et A. Monfort (92) : "Testing Encompassing and Simulating Dynamic Econometric Models ", *Mimeo* No 9214, Crest-Cepremap-Insee.

[40] Gourieroux C., A. Monfort et E. Renault (92) : "Indirect Inference " *Mimeo*, Crest-Cepremap, Insee, Gremaq.

[41] Gourieroux C., A. Monfort et C. Tenreiro (94) : "Kernel M-Estimators : Nonparametric diagnostics for structural models ", *Mimeo* No 9405, Crest-Cepremap-Insee.

[42] Gourieroux C., A. Monfort et A. Trognon (83) : " Testing nested or non-nested hypotheses" *Journal of Econometrics*, Vol.21.

[43] Govaerts B., D. Hendry and J.F. Richard (94) : "Encompassing in Stationary Linear Dynamic Models", *Journals of Econometrics*, Vol. 63, pp. 245-270.

[44] Györfi L., W. Härdle, P. Sarda and P. Vieu (89) : "Nonparametric Curve Estimation from Time Series", *Springer Verlag.*

[45] Hall P. (84) : " Integrated Square Error Properties of Kernel Estimators of Regression Functions ", *Annals of Statistics,* Vol. 12, no 1, pp.241-260.

[46] Härdle W. (90) : " Applied nonparametric regression ". *Econometric society monographs*, Cambridge University Press.

[47] Härdle W., P. Hall and J.S. Marron (92) : "Regression smoothing that are not far from their optimum", *Journal of the Royal Statistical Society*, Series B, Vol.87, no. 417, pp. 227-233.

[48] Härdle W. and S. Luckhaus (84) : "Uniform Concistency of a Class of Regression Function Estimators", *Annals of Statistics,* Vol. 12, no 2, pp. 612-623.

[49] Härdle W. and E. Mammen (93) : "Comparing nonparametric versus parametric regression fits", *Annals of Statistics,* Vol. 21, no 4, pp. 1926-1947.

[50] Härdle W. and J.S. Marron (89) : "Optimal bandwidth selection in nonparametric procedure regression function estimation ", *Annals of statistics*, Vol.13, No.4.

[51] Härdle W. and I. Proença (94) : "Testing a parametric Model Against a Semiparametric Alternative : Practical Considerations ", *Mimeo,* Humboldt University of Berlin.

[52] Hausman J.A. (78) : " Specification test in Econometrics", *Econometrica*, vol 46, no 6.

[53] Hendry D. (93) : " The Roles of Economic Theory and Econometrics in Time Series Economics" , Mimeo, Nuffield College, Oxford.

[54] Hendry D. and J.F. Richard (89) : " Recent development in the theory of encompassing " in *Contribution to operation research and economics*, MIT Press.

[55] Huber P.J. (67) : "The Behavior of Maximum Likelihood Estimates under Non-Standard Conditions", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* no.1, Berkeley, The University of California Press.

[56] Jansen P., R. Serfling and N. Veraverbeke (84) : " Asymptotic Normality for a General Class of Statistical Functions and Applications to Measures of Spread", *Annals of Statistics,* Vol. 12, no 4, pp.1369-1379.

[57] Kulback S and R. A. Leibler (51) : " On information and sufficiency ", *Annals of Mathematical Statistics,* no.22, pp. 79-86.

[58] Lavergne P. (93) : "Sélection non-paramétrique de régresseurs", *Thèse de Doctorat,* Université des Sciences Sociales de Toulouse.

[59] Lavergne P. et Q.H. Vuong (94) : "Nonparametric selection of regressors : The nonnested case", *Mimeo* INRA, no. 94-04D.

[60] Le Cam L. (64) : "Sufficiency and Approximate Sufficiency", *Annals of Mathematical Statistics,* Vol. 35, pp. 1419-1455.

[61] Lu M. and G. E. Mizon (93) : " The Encompassing Principle and Specification Test", *Mimeo,* Economics Departement, Southampton University, UK.

[62] Mack Y.P. and H.G. Müller (89) : "Convolution type estimators for nonparametric regression", *Statistic and Probability Letters,* Vol. 7, pp. 229-239.

[63] Métivier M. (68) : " Notions fondamentales de la théorie des probabilités", Seconde édition, Dunod, Paris.

[64] Marron J. S. (88) : "Automatic Smoothing Parameter Selection : A Survey", Empirical Economics, Vol. 13, pp. 187-208.

[65] Mizon G. E. (84): " The Encompassing Approach in Econometrics " In *Econometrics and Quantitative Economics,* edited by D. F. Hendry and K. F. Wallis. Ch. 6 Oxford : Basil & Blackwell

[66] Mizon G. E. and J. F. Richard (86) : " The encompassing principle and its application to testing non-nested hypotheses " *Econometrica*, Vol.54, No 3.

[67] Nadaraya E.A. (64) : "On estimating regression", *Theory of Probability and its Applications*, 9, pp. 141-142.

[68] Nolland D. and D. Pollard (87) : "U-Processes : Rates of Convergence", *Annals of Statistics,* Vol. 15, no 2, pp. 780-799.

[69] Nychka D. (91) : "Choosing a Range for the Amout of Smoothing in Nonparametric Regression ", *Journal of the Royal Statistical Society*, Series B, Vol.86, no. 415, pp. 653-664.

[70] Pesaran M.H. (74) : "On the General Problem of Model Selection", *Review of Economic Studies*, no 41, pp. 153-171.

[71] Portier B. (92) : "Estimation Non-Paramétrique et Commande Adaptative de Processus Markoviens Non Linéaires", *Thèse de Doctorat,* Centre d'Orsay, Université de Paris Sud.

[72] Rao P. (83) : " Nonparametric functional estimation ", A Probability and mathematical statistics monographs and textbooks, *Academic press.*

[73] Rice J. (84) : "Bandwidth Choice for Nonparametric Regression" , *Annals of Statistics,* Vol. 12, no 4, pp. 1215-1230.

[74] Robinson P. M. (83) : "Nonparametric estimators for Time Series", *Journal of Time Series Analysis*, Vol. 4, No. 3, pp. 185-207.

[75] Rutherford B. and S. Yakowitz (91) : "Error Inference for Nonparametric Regression", *Annals of the Institute of Statistical Mathematicss,* Vol 43, no 1, pp. 115-129.

[76] Sarda P. et P. Vieu (88) : "Vitesses de convergence d'estimateurs nonparamétriques d'une régression et de ses dérivées", *Comptes-rendus de l'Académie des Sciences de Paris*, tome 306, Séries 1, pp. 393-415.

[77] Sawa T. (78) : "Information Criteria for Discriminating Among Alternative Regression Models", *Econometrica,* no. 46, pp. 1273-92.

[78] Schuster E.F. (72) : "Joint asymptotic distribution of the estimated regression function at a finite number of distinct points", *Annals of Mathematical Statistics,* no *43*, pp. 84-88.

[79] Serfling (80) : "Approximation theorems ", *Wiley series in probability and mathematical statistics,* Wiley & Sons.

[80] Shreider Y.A. (66) : "The Monte Carlo Method", *International series of monographs in pure and applied mathematics*, Vol.87, Pergamon Press.

[81] Silverman B. W. (78) : "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives",*Annals of Statistics,* Vol. 6, no 1, pp. 177-184.

[82] Silverman B. W. (85) : "Some aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve Fitting", *Journal of the Royal Statistical Society*, series B, Vol. 47, no. 1, pp. 1–52.

[83] Stone, C. J. (82) : "Optimal global rates of convergence for nonparametric regression", *Annals of Statistics,* Vol 10, no 4, pp. 1040-1053.

[84] Ullah A. (86) : "Nonparametric Estimation and Hypothesis Testing in Econometric Models", *Empirical Economics,* Vol. 13, pp. 223-249.

[85] Ullah A. and H. D. Vinod (93) : "General Nonparametric Regression Estimation and Testing in Econometrics", in *Handbooks of Statistics,* Vol.11, G. S. Maddala, C. Rao and H. D. Vinod editors, Elsevier Science Publishers.

[86] Vieu P. (91) : "Nonparametric Regression : Optimal Local Bandwidth Choice", *Journal of the Royal Statistical Society*, series B, no 53, pp. 453-464.

[87] Vieu P. (93) : "Bandwidth selection for kernel regression", *Computational Statistics and Data Analysis*, à paraître.

[88] Youndje E. (92) : "Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle", *Publication de l'Université de Mont-Saint-Aignan*, No 92-03.

[89] Watson, G.S. (64) : "Smooth regression analysis", *Sankhya, Series A*, 26, pp. 359-372.

[90] White H. (80) : "Using least squares to approximate unknown regression function " *International Economic Review*, Vol.21(1).

[91] White H. (82) : "Maximum likehood estimator of misspecified models " *Econometrica*, Vol.50, pp. 1–26.

[92] White H. (84) : "Asymptotic Theory for Econometricians", *A Series of Monographs and Textbooks*, Academic Press.

[93] Zellner A. (71) : "An Introduction to Bayésian Inférence in Econometrics", *Willey & Sons*, New York.