

Les notions d'enveloppement

Christophe Bontemps*

GREMAQ, Université de Toulouse

Janvier 1995

1 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

*Je remercie Pascal Lavergne pour ses commentaires sur la première version de ce document issu de ma thèse.

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 [32], Mizon et Richard [33] et Hendry et Richard [26]), et de ceux de Gourieroux, Monfort et Trognon, d'autre part (voir Gourieroux et Monfort [18] et [19] ainsi que Gourieroux, Monfort et Trognon [22]). L'extension de ces travaux au cadre bayésien, relié à la notion de spécificité (voir Florens, Hendry et Richard [14]), 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 [22] dans un contexte dynamique a permis l'introduction d'une procédure opérationnelle d'information indirecte [20]. L'ensemble de ces auteurs préconise également l'emploi de tests basés sur ce principe, et plus particulièrement Hendry [25].

L'étude de l'enveloppement est l'objet de cet article, où nous discuterons des définitions précises formalisant ce principe, toutefois une brève discussion informelle clarifie l'analyse.

1.1 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 [14], 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

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

probabilités $\mathcal{P} = \{P_\theta, \theta \in \Theta_0\}$ sur l'espace mesurable (Ω, \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 $(\mathcal{M}, \hat{\delta})$, au lieu de \mathcal{M} , toutefois, après avoir levé toute ambiguïté, nous ignorerons cette notation.

On cherche à confronter un modèle $(\mathcal{M}_1, \hat{\beta})$ avec un modèle rival $(\mathcal{M}_2, \hat{\gamma})$, où $\hat{\beta}$ et $\hat{\gamma}$ sont deux estimateurs convergents des paramètres β et γ respectivement, appartenant aux espaces paramétriques, 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 $\hat{\gamma}$ à partir de $\hat{\beta}$, 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 [19] présente l'originalité de supposer le processus de génération des données extérieur aux modèles en présence. Cet article 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 concurrents². 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 construc-

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

tivement, 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 [26] dans cette voie, où l'enveloppement relève plus de la *comparaison* de modèles que du *choix* de modèles.

1.2 Enveloppement exact ou approché ?

L'enveloppement (“*exact*”), 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 [28](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 [35], la pseudo-vraie valeur est définie comme l'élément (s'il existe) minimisant le KLIC. Cette définition, qui figure également chez White [37] ou Gourieroux, Monfort et Trognon [22], semble avoir été introduite (implicitement) dans l'oeuvre de Cox [6] et [7] relative aux tests d'hypothèses non-emboîtées, ainsi que dans les travaux de Huber [27].

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

1.3 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 5). 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 Chib et Greenberg [8]), ou par des techniques d'approximation qui permettent un calcul opérationnel (voir Florens, Hendry et Richard [14]). Malheureusement, ces procédures ne sont encore définies que pour des cas particuliers (voir Florens, Larribeau et Mouchart [16]).

1.4 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 [24] et White [38] 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 [22] 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 la pratique l'on utilise souvent de modèles (et donc d'estimateurs) éventuellement fonctionnels, l'approche asymptotique sera bien évidemment privilégiée.

1.5 Emboîtés ou non-emboîtés ?

Dans son article sur le problème général de la sélection de modèles, Pesaran [34] 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 ([6] et [7]), 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 [34]).

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 [?], 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 [10], 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 séquant 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 .

Hendry et Richard [26] 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 3.5, section 4.1), 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 4.1), 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*”.

Cet article se veut une clarification de la notion d’enveloppement. Dans la section 2 nous définissons l’enveloppement exact et en caractérisons les propriétés dans un cadre statique et dynamique. Cette notion sera distinguée de celle d’enveloppement approché dans la section 3, basée sur les pseudo-vraies valeurs à distance finie ou asymptotiques. Nous examinerons ensuite l’enveloppement parcimonieux et partiel en section 4 ainsi que l’enveloppement Bayésien section 5. La section 6 conclue le papier en présentant quelques tests. ????????

2 Définition de l'enveloppement exact

‘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)

Soit Y une variable aléatoire définie sur l'espace mesurable (Ω, \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 $\hat{\beta}_n$ et $\hat{\gamma}_n$ des estimateurs consistants de β et γ dans leurs modèles respectifs, les estimateurs $\hat{\beta}_n$ et $\hat{\gamma}_n$ 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 [26] :

Définition 2.1 (*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 :

$$\hat{\gamma}(Y_n) = \Gamma(\hat{\beta}(Y_n)) \quad (M_1 \text{ p.s.}) \quad (1)$$

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 $\hat{\gamma}(Y_n)$ à partir de $\hat{\beta}(Y_n)$. \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 \mathfrak{R}^+ et représentés par les densités suivantes:

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

munis des estimateurs

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

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 $\hat{\gamma} = \Gamma(\hat{\beta})$.

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

Exemple 2:

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

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

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

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

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

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

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

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

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

$$\hat{\gamma} = \begin{pmatrix} \hat{\eta} \\ 1 \end{pmatrix} = \begin{pmatrix} \bar{y}^{(1)} \\ 1 \end{pmatrix}$$

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

$$\Gamma : \begin{array}{ccc} \mathfrak{R}^2 & \longrightarrow & \mathfrak{R}^2 \\ \begin{pmatrix} u \\ v \end{pmatrix} & \longrightarrow & \begin{pmatrix} u \\ 1 \end{pmatrix} \end{array}$$

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 4, 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 $\tilde{\gamma} \neq \hat{\gamma}$. 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 $\tilde{\gamma}$ estimateur du maximum de vraisemblance :

$$\tilde{\gamma} = \begin{pmatrix} \tilde{\eta} \\ 1 \end{pmatrix} \quad \text{avec} \quad \tilde{\eta} = \bar{y}^1 + \pi(\bar{y}^2 - 1)$$

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

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

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

Nous avons, ici également, $\tilde{\gamma} = \Gamma'(\hat{\beta})$, 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 $(\mathcal{M}_1, \hat{\beta})$ enveloppe le modèle $(\mathcal{M}_2, \hat{\gamma})$ ainsi que le modèle $(\mathcal{M}_2, \tilde{\gamma})$. \square

2.1 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 :

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

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

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

Govaerts, Hendry et Richard [23], proposent la définition de l'enveloppement dynamique, dans le même esprit que la définition 2.1 :

Définition 2.2 : "Le modèle dynamique \mathcal{M}_1 enveloppe exactement \mathcal{M}_2 ", s'il existe une séquence de fonctions Γ_T telle que :

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

Ici encore, et pour tout T , la connaissance de $\hat{\beta}_T$ associée à celle des fonctions de lien Γ_T , permet la connaissance de l'estimateur de \mathcal{M}_2 , $\hat{\gamma}_T$. 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) que par l'aspect séquentiel que doit revêtir ici la fonction de lien Γ , remplacée ici par une succession de fonctions de liens.

2.2 Propriétés

Nous pouvons reformuler la définition 2.1 d'une manière plus visuelle en examinant les relations entre les espaces Ω , Θ_β et Θ_γ :

Définition 2.1 (bis) :

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

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

Nous verrons section 5 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) 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).

2.3 Transitivité de l'enveloppement exact

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

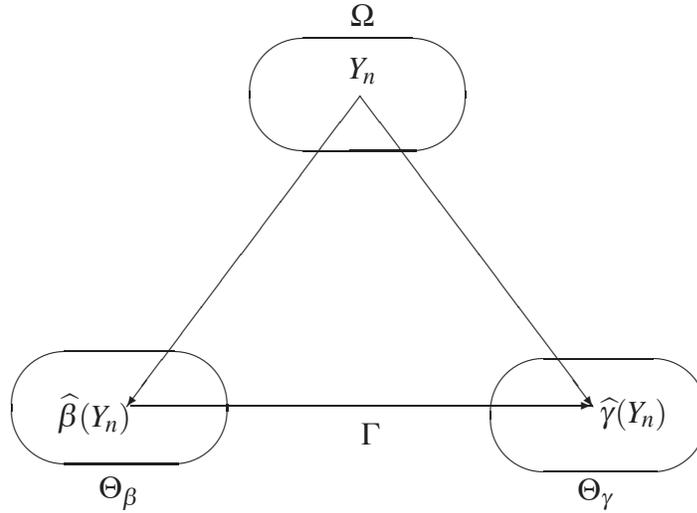


Figure 1: Enveloppement exact

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

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

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

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

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

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

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

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

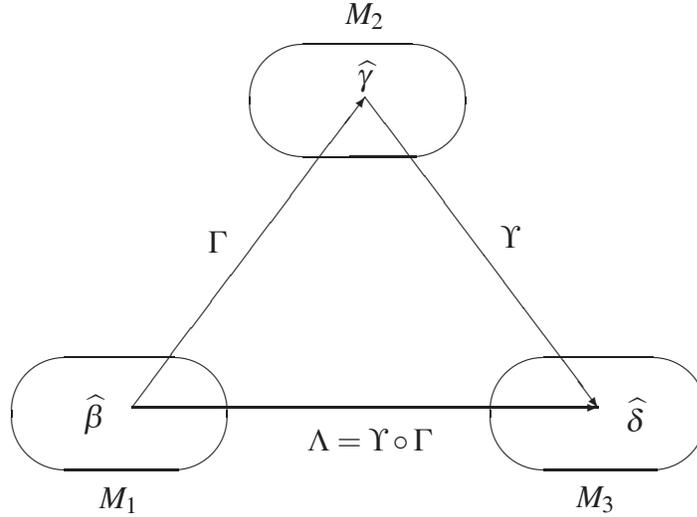


Figure 2: Transitivité de l’enveloppement exact

et donc : $\{Y_n \text{ tels que } \hat{\gamma}(Y_n) \neq \Gamma(\hat{\beta}(Y_n))\}$ est de mesure nulle pour \mathcal{M}_1 .

D’autre part :

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

c’est-à-dire que l’ensemble : $\{Y_n \text{ tels que } \hat{\delta}(Y_n) \neq \Upsilon(\hat{\gamma}(Y_n))\}$ est de mesure nulle pour \mathcal{M}_2 .

L’égalité : $\hat{\delta} = \Lambda(\hat{\beta}) = \Upsilon \circ \Gamma(\hat{\beta}) \quad \mathcal{M}_1 \text{ presque sûrement}$

ne sera vérifiée que si l’ensemble $\{Y_n \text{ tels que } \hat{\delta}(Y_n) \neq \Upsilon(\hat{\gamma}(Y_n))\}$ 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.

3 Enveloppement approché

D’une manière générale, l’enveloppement exact défini en (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.

3.1 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 $\Psi(\hat{\gamma}, \hat{\beta})$ mesurant l’écart, ou la divergence, entre les modèles \mathcal{M}_1 et \mathcal{M}_2 .

La fonction de lien Γ

$$\Gamma : \begin{array}{l} \Theta_\beta \longrightarrow \Theta_\gamma \\ \hat{\beta}(Y_n) \longrightarrow \Gamma(\hat{\beta}(Y_n)) \end{array}$$

qui détermine la pseudo-vraie valeur $\Gamma(\hat{\beta}(Y_n))$, 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.

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

Cette fonction Γ minimise la “spécificité” de \mathcal{M}_2 vis à vis de \mathcal{M}_1 , c’est égale-

ment celle qui donne le plus de possibilités à \mathcal{M}_1 d'expliquer les résultats de \mathcal{M}_2 ³. 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

Dans les modèles paramétriques de maximum de vraisemblance, le critère d'information de Kulback-Leibler [28] est généralement adopté comme “*distance*” entre modèles⁴. Nous vérifierons que ce critère coïncide avec une mesure de la spécificité introduite par Florens, Hendry et Richard [14].

3.2 Le contraste de Kulback et Leibler (KLIC)

Soit $(\mathfrak{X}, \mathcal{B}_{\mathfrak{X}}, \lambda)$ 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 .

³La “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 :

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

Cette spécificité est évidemment conditionnelle à l'échantillon 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.

⁴Ce critère porte souvent le nom de “*contraste*”, exprimant ainsi l'idée qu'il s'agit d'un 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.

- 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 $\beta \in \Theta_\beta \subset \mathfrak{R}^l$
- ii) Le modèle \mathcal{M}_1 suppose que Y admet la densité $g(y, \gamma)$ par rapport à λ , avec g continue en $\gamma \in \Theta_\gamma \subset \mathfrak{R}^m$
- 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 [17].

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

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

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, Hendry et Richard [14] proposent de définir la pseudo-vraie valeur comme minimisant la *spécificité* de $(\mathcal{M}_2, \hat{\gamma})$ vis-à-vis de $(\mathcal{M}_1, \hat{\beta})$.

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

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

Cette mesure est évidemment dépendante de l’échantillon Y_n ⁵.

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

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

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 $\hat{\gamma}$ est l’estimateur du maximum de vraisemblance, on a :

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

⁵Une 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 :

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

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 5). La mesure précédente est cependant préférée, elle conduit en effet à une présentation naturelle des distances entre modèles.

3.3 Définition de la pseudo-vraie valeur

Gourieroux, Monfort et Trognon [22], sur les bases des travaux de Sawa [35], Huber [27] et Cox ([6] et [7]), 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 :

$$\Gamma(\beta) = \arg \max_{\delta} \int_{\Omega} \log [g(y, \delta)] f(y, \beta) \lambda(dy) \tag{2}$$

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 :

$$\Gamma(\beta) = E_{\beta}(\hat{\gamma}) \quad (3)$$

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

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

$$\Gamma(\beta) = \int_{\Omega} \arg \max_{\delta} \log [g(y, \delta)] f(y, \beta) \lambda(dy) \quad (4)$$

La distinction entre (2) et (4), réside alors dans l'ordre des opérateurs. Conformément à Hendry, Mizon et Richard (Voir Mizon [32], Mizon et Richard [33], ou Hendry et Richard [26]), l'espérance sous \mathcal{M}_1 de $\hat{\gamma}$ définissant la pseudo-vraie valeur dans l'expression 3 est remplacée par :

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

$\Gamma(\beta)$ se présente ici comme une réinterprétation de l'estimateur $\hat{\gamma}$ par \mathcal{M}_1 , elle est aisément estimable, dès lors que β l'est, par $\Gamma(\hat{\beta})$. Nous utiliserons cette expression de la pseudo-vraie valeur dans la suite de ce travail. Gourieroux et alii [22], proposent également une définition de la pseudo-vraie valeur en échantillon fini dont $\Gamma(\beta)$ est la limite.

3.4 Pseudo-vraie valeur à distance finie

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

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

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

Les log-vraisemblances conditionnelles associées à ces modèles sont⁶:

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

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

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

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 :

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

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 :

$$\sum_{i=1}^n E_\beta [\log g(y_i | x_i, \gamma)] \tag{5}$$

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

$$\max_{\gamma \in \Theta_\gamma} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E_\beta [\log g(y_i | x_i, \gamma)] = \max_{\gamma \in \Theta_\gamma} E_x E_\beta [\log g(y_i | x_i, \gamma)] \tag{6}$$

Où E_x désigne l'espérance relative à la distribution des x_i ⁷.

⁶Les 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 [17], volume 2, ou dans l'étude des pseudo-vraies valeurs réalisée par Dhaene [11].

⁷On supposera, en effet, que la distribution empirique des x_i tend vers une distribution limite (et inconnue).

Remarque

La pseudo-vraie valeur à distance finie $\Gamma_n(\beta)$, maximum de l'expression (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 (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.

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 $\hat{\gamma} - \Gamma(\hat{\beta})$. L'expression de cette différence est centrale dans cette définition, elle servira de base aux tests d'enveloppement. 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) de la définition 2.1.

3.5 Définition de l'enveloppement approché

Comme il n'est pas possible de vérifier la relation (1), l'idée est de définir l'enveloppement approché en se basant sur la différence entre l'estimateur $\hat{\gamma}$ de γ dans \mathcal{M}_2 et un estimateur $\Gamma(\hat{\beta})$ de la pseudo-vraie valeur $\Gamma(\beta)$, celle-ci ayant été calculée par minimisation de la spécificité.

Définition 3.1 (*Enveloppement approché*) :

On dira que " \mathcal{M}_1 enveloppe \mathcal{M}_2 " ($\mathcal{M}_1 \mathcal{E} \mathcal{M}_2$) si :

$$\hat{\gamma}(Y_n) = \Gamma(\hat{\beta}(Y_n)) \quad (M_1 \text{ p.s.}) \quad (7)$$

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

La différence fondamentale avec l'expression (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 $\hat{\gamma} - \Gamma(\hat{\beta})$, 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 (7) est évidemment dépendante de l'échantillon, et peut donc être testée. C'est d'ailleurs sur la différence $\hat{\gamma} - \Gamma(\hat{\beta})$, ou sur une fonction de cette différence que seront fondés les tests d'enveloppement classiques (voir Bontemps [3] pour une revue).

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 [11]).

L'exemple suivant permet de représenter l'enveloppement approché sous une forme aussi simple que possible. Il est extrait de Hendry et Richard [26] 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 $Y_n = (y_1, y_2, \dots, y_n)$, de n réalisations indépendantes de la variable aléatoire réelle Y .

Les estimateurs associés aux paramètres de ces modèles sont :

- $\hat{\beta} = \frac{1}{n} \sum_{i=1}^n y_i$, 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 $\Gamma(\beta) = p \lim_{\mathcal{M}_1} (\widehat{\gamma}^2) = 1 + \beta^2$

\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 :

$$\widehat{\phi} = \widehat{\gamma}^2 - \Gamma(\widehat{\beta}) = \widehat{\gamma}^2 - 1 - \widehat{\beta}^2 \quad (8)$$

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 $\widehat{\phi} = \widehat{\beta}$ est nul . □

3.6 L'alternative de Gourieroux et Monfort

L'approche de Gourieroux et Monfort [18], [19] et [22], 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 3.4. Deux modèles sont proposés pour la modélisation de la densité conditionnelle de y sachant x .

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

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

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 :

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

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 :

$$\begin{aligned}
\beta_0 &= \text{Arg min}_{\beta} E_x E_0 \log \left[\frac{h(y_i | x_i, \theta_0)}{f(y_i | x_i, \beta)} \right] \\
&= \text{Arg max}_{\beta} E_x E_0 [\log f(y_i | x_i, \beta)] \\
\text{et} \\
\gamma_0 &= \text{Arg max}_{\gamma} E_x E_0 [\log g(y_i | x_i, \gamma)]
\end{aligned}$$

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 :

$$\Gamma(\beta) = \text{Arg max}_{\gamma} E_x E_{\beta} \log g(y_i | x_i, \gamma)$$

La fonction déterminant $\Gamma(\beta)$ dans Θ_{γ} , est la "fonction de lien" $\Gamma : \Theta_{\beta} \rightarrow \Theta_{\gamma}$, définie section 3.3.

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

$$\mathcal{B}(\gamma) = \text{Arg max}_{\beta} E_x E_{\gamma} \log f(y_i | x_i, \beta)$$

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 :

$$\begin{aligned}
\mathcal{M}_1 &: f(y | x, \beta) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(y - x'_1 \beta)^2}{2} \right) \\
\text{et} \\
\mathcal{M}_2 &: g(y | x, \gamma) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{(y - x'_2 \gamma)^2}{2} \right)
\end{aligned}$$

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

$$\begin{aligned}
\Gamma(\beta) &= \text{Arg max}_{\gamma} E_x E_{\beta} [\log g(y | x, \gamma)] \\
&= \text{Arg max}_{\gamma} E_x E_{\beta} - (y - x'_2 \gamma)^2 \\
&= \text{Arg max}_{\gamma} E_x (x'_1 \beta - x'_2 \gamma)^2 \\
&= (E_x [x_2 x'_2])^{-1} E_x [x_2 x'_1] \beta
\end{aligned}$$

De même :

$$\begin{aligned}
\mathcal{B}(\gamma) &= \text{Arg max}_{\beta} E_x E_{\beta} [\log g(y | x, \gamma)] \\
&= (E_x [x_1 x'_1])^{-1} E_x [x_1 x'_2] \gamma
\end{aligned}$$

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 $\hat{\beta}$ et $\hat{\gamma}$ dans les expressions ci-dessus. \square

Dans ce contexte conditionnel, Gourieroux et Monfort [22] nous donnent leur définition de l'enveloppement, également proposée par Hendry et Richard [26], sous le terme “d’enveloppement global” (“population encompassing”).

Définition 3.2 : \mathcal{M}_1 enveloppe \mathcal{M}_2 sous \mathcal{P}_0 ssi :

$$\gamma_0 = \Gamma(\beta_0) \tag{9}$$

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 (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 $\forall \mathcal{M}_2$:

$$\begin{aligned}\gamma_0 &= \text{Arg max}_{\gamma} E_x E_0 [\log g(y_i | x_i, \gamma)] \\ &= \text{Arg max}_{\gamma} E_x E_{\mathcal{M}_1} [\log g(y_i | x_i, \gamma)] \\ &= \Gamma(\beta_0)\end{aligned}$$

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 est⁸ :

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

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

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

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 :

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

et

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

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 :

⁸ \mathcal{M}_1 peut être défini également comme $\{f(y_i | x_i, \beta) \text{ t.q. } \beta \in \Theta_{\beta}\}$

$\mathcal{M}'_2 = \text{Im}(\mathcal{M}_1)$ est le sous ensemble de \mathcal{M}_2 dont les paramètres γ appartiennent à l'image de la matrice $(E_x[x_2x'_2])^{-1} E_x[x_2x'_1]$. \square

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 [19]), 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 (3.4). Gourieroux et Monfort [20] proposent également une procédure de simulation de ces pseudo-vraies valeurs finies par tirages aléatoires d'éléments observés du processus (y_i, x_i) .

4 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)

4.1 Enveloppement parcimonieux

Définition 4.1 : \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 ⁹

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). 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 [31].

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 3.5, 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.

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

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

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

⁹L'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.

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

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

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

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 :

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

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

\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 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 . \square

4.2 Enveloppement partiel

Lu et Mizon[31] 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 4.2 (*Enveloppement via une fonction*) :

$$\mathcal{M}_1 \text{ enveloppe } \mathcal{M}_2 \text{ via } C \quad \text{ssi} \quad C(\gamma) - C(\Gamma(\beta)) = 0 \quad (10)$$

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 4.2-bis (Lu et Mizon) :

$$\mathcal{M}_1 \text{ enveloppe } \mathcal{M}_2 \text{ via } B \quad \text{ssi} \quad E_{\mathcal{P}_0} [B(Y_n, \hat{\gamma}) - E_{\mathcal{M}_1} [B(Y_n, \hat{\gamma})]] = 0 \quad (11)$$

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 $\hat{\gamma}$.

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

- L'estimateur $\hat{\gamma}$ est en fait généralisé par la fonction $B(Y_n, \hat{\gamma})$ dans l'expression (11),
- tandis que $E_{\mathcal{M}_1} [B(Y_n, \hat{\gamma})]$, qui est la réinterprétation de $B(Y_n, \hat{\gamma})$ sous \mathcal{M}_1 , remplace $\Gamma(\beta) = E_{\beta} [\hat{\gamma}]$, la pseudo-vraie valeur de $\hat{\gamma}$.

La différence est toutefois analysée ici sous \mathcal{P}_0 . Cette définition trouve sa source dans l'article de Mizon et Richard [33], 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 :

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

où $L_1(\hat{\gamma})$ et $L_2(\hat{\beta})$ 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 ([6] et [7]), que l’on trouve clairement explicité dans Pesaran [34]. 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 [32], Mizon et Richard [33] et Lu et Mizon [31]. 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 :

$$\mathcal{M}_1 \text{ enveloppe exactement } \mathcal{M}_2 \implies \mathcal{M}_1 \text{ enveloppe } \mathcal{M}_2 \text{ via } C$$

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 $\hat{\gamma}$ (une projection par exemple), seule “une partie” des paramètres de \mathcal{M}_2 est alors considérée comme pertinente pour l’analyse. Dans sa thèse, Bontemps [3] montre 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 .

5 Enveloppement bayésien

“In summary, encompassing is formalized as a concept of sufficiency among models whereas specificity measures the lack of encompassing”

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

5.1 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 modèles¹⁰. 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 (12). Nous rappelons ici cette décomposition pour le modèle \mathcal{M}_1 :

$$\begin{aligned} \pi(y, \beta) &= f(y | \beta) \cdot \mu(\beta) \\ &= \mu(\beta | y) \cdot f(y) \end{aligned} \tag{12}$$

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

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

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.

5.2 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*

¹⁰Nous supposons, dans toute cette partie, que les modèles bayésiens sont représentés par des densités, sans donner les conditions nécessaires à cette propriété.

posteriori de \mathcal{M}_2 en utilisant celle de \mathcal{M}_1 . Hendry et Richard [26] proposent une comparaison de l'a posteriori bayésien de \mathcal{M}_2 avec une interprétation de cet a posteriori par le modélisateur 1.

Définition 5.1 (*Enveloppement bayésien*) :

On dira que “le modèle bayésien \mathcal{M}_1 enveloppe \mathcal{M}_2 ”, s’il existe une densité conditionnelle $\Gamma(\gamma | \beta)$ indépendante de y , telle que :

$$v(\gamma | y) = \int_{\Theta_\beta} \Gamma(\gamma | \beta) \mu(\beta | y) \partial\beta \quad (13)$$

“presque sûrement ” en y .¹¹

L’expression (13) exprime le fait que les résultats de \mathcal{M}_2 , c’est-à-dire $v(\gamma | y)$, sont retrouvés à l’aide de ceux de \mathcal{M}_1 , $\mu(\beta | y)$.

Le lien entre l’enveloppement classique et l’enveloppement bayésien réside dans la fonction Γ qui permettrait de lier les espaces paramétriques et qui est remplacée ici par la densité de transition $\Gamma(\gamma | \beta)$ 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 (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 $\Gamma(\gamma | \beta)$. Auparavant, nous suivrons Florens, Hendry et Richard [14], sur la voie de la dualité existant entre l’enveloppement bayésien et l’exhaustivité entre statistiques au sens de Le Cam [30].

5.3 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 :

¹¹Dans une optique bayésienne “presque sûrement ” s’entend ici au sens de la loi prédictive de \mathcal{M}_1 , de densité $f(y)$.

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 :

$$g(z | \beta) = \int_{\mathcal{Y}} f(y | \beta) \Lambda(z | y) \partial y \quad (14)$$

où:

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

Selon Hendry et Richard, l'expression (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 (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 [30]. 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 "*déficiance (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 $\Gamma(\gamma | \beta)$.

5.4 Enveloppement bayésien et spécificité

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

Définition 5.2 Soient (A, \mathcal{A}) et (C, \mathcal{C}) deux espaces mesurables, une "*probabilité de transition*" est une fonction Λ :

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

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 $\Gamma(\gamma | \beta)$ 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 $\Gamma(\gamma | \beta)$ ¹².

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

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 :

$$v^*(\gamma | y) = \int_{\Theta_\beta} \mu(\beta | y) \Gamma(\gamma | \beta) d\beta \quad (15)$$

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 [30], Hendry et Richard [26] 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 , $v^*(\gamma | y)$ et $v(\gamma | y)$. Suivant les mesures choisies pour quantifier cette différence ou divergence (voir Hendry et Richard [26]), on obtiendra la spécificité, la p-spécificité, ou la φ -spécificité, comme minimum de la divergence espérée entre $v^*(\gamma | y)$ et $v(\gamma | y)$. Cette spécificité représente en fait la quantité incompressible séparant \mathcal{M}_1 de \mathcal{M}_2 . Cette

¹²Florens et alii [14] 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 :

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

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

notion de divergence espérée minimale correspond à celle utilisée par Le Cam dans le contexte dual, pour mesurer le défaut d'exhaustivité entre statistiques.

5.5 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 (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 bayésienne 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 [14]) Des méthodes de simulation sont toutefois capables de déterminer numériquement cette transition optimale, (voir Florens, Larribeau et Mouchart [16]), comme l'échantillonneur de Gibbs. Une autre voie consiste à approcher la pseudo-vraie valeur, et à considérer l'enveloppement approché basé sur cette pseudo-vraie valeur. Dans leur article sur l'enveloppement bayésien, Florens, Hendry et Richard [14], 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 plim 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 , $v(\gamma | y)$, dans l'optique de \mathcal{M}_1 . La pseudo-vraie valeur approchée $\tilde{\Gamma}(\gamma | \beta)$ est donc obtenue comme marginalisation de l'*a posteriori* $v(\gamma | y)$ de \mathcal{M}_2 , par rapport à l'échantillon, en utilisant la densité d'échantillonnage de \mathcal{M}_1 , $f(y | \beta)$:

$$\tilde{\Gamma}(\gamma | \beta) = \int v(\gamma | y) \cdot f(y | \beta) dy$$

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

$$\tilde{\Gamma}(\gamma | \beta) = \frac{1}{\mu(\beta)} \int v(\gamma | y) \cdot \mu(\beta | y) f(y) dy$$

Cette densité de transition, bien que n'étant pas la transition optimale, présente

l'avantage d'être facilement calculable dans un grand nombre d'applications¹³.

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

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

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

qui est solution du système :

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

Si $E_1 \left(\hat{\beta} \hat{\beta}' \right)$ est non singulière alors $\hat{\Lambda}$, (et donc $\hat{\Gamma}(\gamma | \beta)$), 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 $\hat{\gamma}$ converge vers $\gamma(\beta)$ sous la loi jointe de \mathcal{M}_1 alors :

$$\hat{\Lambda} \rightarrow [E_1(\beta\beta')]^{-1} E_1(\beta\gamma'(\beta))$$

- 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 :

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

¹³En effet, dans le cas où les deux modèles présentent les mêmes densités a priori et d'échantillonnage, la densité de transition $\tilde{\Gamma}(\gamma | \beta)$ se retrouve réduite à une Dirac.

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.

6 Conclusion

“An essential characteristic of empirical modelling (and in fact in the development 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 cette différence entre estimateur et pseudo-vraie valeur que vont être fondés les tests d'enveloppement 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 [33], Gourieroux et Monfort [19], et Florens, Hendry et Richard [14] dans un contexte bayésien. L'hypothèse nulle ainsi que le modèle de référence lors de ces tests diffère suivant les auteurs. Dans leur étude, Hendry et Richard [26], 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 [18], 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 $\hat{\phi}_0 = (\hat{\gamma}_0 - \Gamma(\hat{\beta}_0))$ 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 [33], étudient l'enveloppement sous l'optique du modèle \mathcal{M}_1 , l'hypothèse nulle, directement issue de l'enveloppement exact, est alors :

$$H'_0 \quad : \quad \gamma = \Gamma(\beta)$$

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 $\hat{\phi}$ défini comme $\hat{\gamma} - \Gamma(\hat{\beta})$, ou d'une fonction de $\hat{\phi}$. En règle générale la distribution de $\hat{\phi}$ n'est pas connue en échantillon fini, et il est alors nécessaire d'avoir recours à une étude asymptotique pour caractériser la distribution de $\hat{\phi}$. Ce cadre s'étend d'ailleurs sous la même forme, aux estimateurs fonctionnels de la regression pour lesquels des tests paramétriques et non paramétriques ont été développés (voir Bontemps, Florens et Richard [4]).

References

- [1] Amemiya T. (80) : “Selection of regressors”, *International Economic Review*, 21(2), pp. 331-354.
- [2] Amemiya T. (85) : “Advanced Econometrics”, *Basil Blackwell*, Oxford.
- [3] Bontemps, C. (1995) : “Enveloppement dans les modèles de régression paramétriques et non-paramétriques”, *Thèse de Doctorat*, Université des Sciences Sociales de Toulouse.
- [4] Bontemps, C., J. P. Florens et J.F. Richard (1997) : “Encompassing in regression models : Parametric and non-parametric procedures" to appear G. Mizon and M. Marcellino (Eds.) "*Encompassing and non-nested tests*", Oxford University Press. .
- [5] Bouoiyour J. (93) : “ Tests Bayesiens d’enveloppement et fonction de coût”, *Thèse de Doctorat*, Université des Sciences Sociales de Toulouse.
- [6] 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.
- [7] 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.
- [8] Chib Siddhartha abd Edward Greenberg (96):: “ Markov Chain Monte Carlo Simulation Methods in Econometrics” *Econometric Theory*, Vol. 12, No. 3, pp. 409-431.
- [9] 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.
- [10] Davidson R. et J.G. Mac Kinnon (93) : “ Estimation and inference in econometrics ”, *Oxford University Press*.
- [11] Dhaene G. (93) : “ Encompassing : Formulation, properties and testing ” Ph. D Dissertation Université Catholique de Louvain, Louvain-la-Neuve.
- [12] Dhaene G. (93) : “ On the Encompassing Relation ”, *Mimeo*, Université Catholique de Louvain, Louvain-la-Neuve.

- [13] Engle R.F. , D.F.Hendry and J.F. Richard (83) : “Exogeneity”, *Econometrica*, No. 55, pp 277-304.
- [14] Florens J.P., D.F. Hendry and J. F. Richard (96) : “ Encompassing and Specificity ” *Econometric Theory*, Vol. 12, No. 4, pp. 620-656 .
- [15] 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.
- [16] Florens J.P., S. Larribeau et M. Mouchart (94) : “ Bayesian encompassing tests of a unit root hypothesis.” *Econometric Theory* Vol. 10, No. 3/4, Symposium Double Issue: Bayes Methods and Unit Roots , pp. 747-763
- [17] Gouriéroux C. et A. Monfort (89) : “Statistique et modèles économétriques”, Vol. 1 et 2, *Economica*, Paris.
- [18] Gouriéroux C. et A. Monfort (91) : “Testing Non Nested Hypotheses ”, *Mimeo* No 9207, Crest-Cepremap-Insee.
- [19] Gouriéroux C. et A. Monfort (95) : “Testing Encompassing and Simulating Dynamic Econometric Models ” *Econometric Theory*, Vol. 11, No. 2 , pp. 195-228
- [20] Gouriéroux C., A. Monfort et E. Renault (92) : “Indirect Inference ” *Mimeo*, Crest-Cepremap, Insee, Gremaq.
- [21] Gouriéroux C., A. Monfort et C. Tenreiro (94) : “Kernel M-Estimators : Nonparametric diagnostics for structural models ”, *Mimeo* No 9405, Crest-Cepremap-Insee.
- [22] Gouriéroux C., A. Monfort et A. Trognon (83) : “ Testing nested or non-nested hypotheses” *Journal of Econometrics*, Vol.21.
- [23] Govaerts B., D. Hendry and J.F. Richard (94) : “Encompassing in Stationary Linear Dynamic Models”, *Journals of Econometrics*, Vol. 63, pp. 245-270.
- [24] Hausman J.A. (78) : “ Specification test in Econometrics”, *Econometrica*, vol 46, no 6.
- [25] Hendry D. (93) : “ The Roles of Economic Theory and Econometrics in Time Series Economics” , *Mimeo*, Nuffield College, Oxford.

- [26] Hendry D. and J.F. Richard (89) : “ Recent development in the theory of encompassing ” in *Contribution to operation research and economics*, MIT Press.
- [27] 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.
- [28] Kulback S and R. A. Leibler (51) : “ On information and sufficiency ”, *Annals of Mathematical Statistics*, no.22, pp. 79-86.
- [29] Lavergne P. (93) : “Sélection non-paramétrique de régresseurs”, *Thèse de Doctorat*, Université des Sciences Sociales de Toulouse.
- [30] Le Cam L. (64) : “Sufficiency and Approximate Sufficiency”, *Annals of Mathematical Statistics*, Vol. 35, pp. 1419-1455.
- [31] Lu M. and G. E. Mizon (96) : “ The Encompassing Principle and Specification Test”, *Econometric Theory*, Vol. 12, No 5, pp.845-858.
- [32] 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
- [33] Mizon G. E. and J. F. Richard (86) : “ The encompassing principle and its application to testing non-nested hypotheses ” *Econometrica*, Vol.54, No 3.
- [34] Pesaran M.H. (74) : “On the General Problem of Model Selection”, *Review of Economic Studies*, no 41, pp. 153-171.
- [35] Sawa T. (78) : “Information Criteria for Discriminating Among Alternative Regression Models”, *Econometrica*, no. 46, pp. 1273-92.
- [36] Shreider Y.A. (66) : “The Monte Carlo Method”, *International series of monographs in pure and applied mathematics*, Vol.87, Pergamon Press.
- [37] White H. (80) : “Using least squares to approximate unknown regression function ” *International Economic Review*, Vol.21(1).
- [38] White H. (82) : “Maximum likelihood estimator of misspecified models ” *Econometrica*, Vol.50, pp. 1–26.

- [39] White H. (84) : “Asymptotic Theory for Econometricians”, *A Series of Monographs and Textbooks*, Academic Press.
- [40] Zellner A. (71) : “An Introduction to Bayesian Inference in Econometrics”, *Wiley & Sons*, New York.