



**HAL**  
open science

# Comparaison de deux méthodes d'estimation sur la base du RICA

Dominique Desbois

► **To cite this version:**

Dominique Desbois. Comparaison de deux méthodes d'estimation sur la base du RICA. 8<sup>e</sup> Journées de Méthodologie Statistique de l'Insee, INSEE, Dec 2002, Paris, France. hal-03319858

**HAL Id: hal-03319858**

**<https://hal.inrae.fr/hal-03319858>**

Submitted on 13 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# COÛTS DE PRODUCTION AGRICOLE :

## Comparaison de deux méthodes d'estimation sur la base du RICA

Dominique DESBOIS

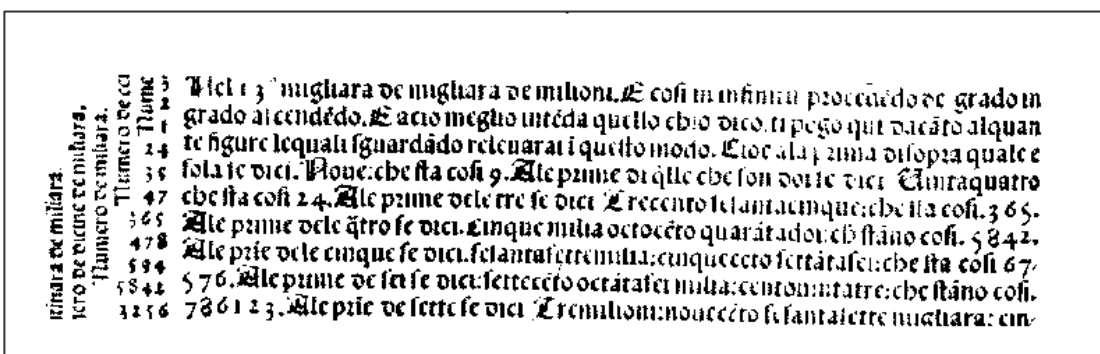
Inra-Economie et Sociologie rurales  
et MAAPAR/DAF/SCEES (Bureau du Rica)



Portrait de Luca Pacioli (1445 – 1514) par Jacopo de Barbari (fragment, 1495).

Ce franciscain originaire de *Sansepolcro* (Italie) fut l'un des mathématiciens les plus influents de son époque non seulement en raison de ses travaux en arithmétique (notamment un calcul approché de la racine carrée qui, comme cas particulier, préfigure la méthode de Newton) mais aussi en tant qu'enseignant, auteur de deux traités de mathématiques : *Summa de arithmetica, geometria, proportioni et proportionalita* (Venise, 1494) et *Divina proportione* (Venise, 1509) dont certaines illustrations sont dues à *Leonardo da Vinci*, qui fût son ami. *Fra Luca Pacioli* est également reconnu comme l'inventeur de la comptabilité en partie double. C'est à ce titre que le séminaire annuel des travaux de recherche en microéconomie agricole réalisés à partir du Réseau d'information comptable agricole (**Rica**) européen porte le nom de Pacioli, en hommage au moine franciscain dont l'œuvre illustre bien l'esprit de la Renaissance.

Fragment de page extraite de l'ouvrage *Summa de arithmetica, geometria proportioni et proportionalita*



# **1. Pour un observatoire des coûts de production agricole**

## **1.1 Les défis majeurs pour l'Europe agricole**

En tant que marché unique, l'Europe est confrontée dans la période actuelle à des défis majeurs au plan économique :

### **1.1.1 Les négociations internationales à l'Organisation mondiale du commerce (OMC)**

Le Cycle du Développement, initié à Doha en novembre 2001, doit aboutir à une plus grande libéralisation des échanges au niveau mondial, donc à une ouverture accrue des marchés européens aux produits des pays tiers, avec en corollaire une diminution des concours publics à l'agriculture européenne. Malgré l'absence d'accord sur le dossier agricole lors de la conférence ministérielle de Cancun, au Mexique en septembre 2003, concilier ouverture des marchés de produits agricoles et maintien de productions viables pour l'ensemble des zones rurales demeure un objectif majeur des négociateurs européens à l'OMC.

### **1.1.2 La réforme de la politique agricole commune (Pac)**

Les réformes successives de la Pac intervenues en 1992 puis en 1999 ont remplacé progressivement le soutien des prix des produits agricoles par un système d'aides directes aux producteurs. L'accord de Luxembourg du 26 juin 2003 entend renforcer le découplage de ces aides par rapport à la production, avec pour objectif de « rendre aux signaux du marché leur place dans les choix productifs des agriculteurs européens ».

### **1.1.3 L'élargissement de l'Union européenne (UE)**

Conformément au traité d'Athènes, dix pays rejoindront le marché unique européen au 1<sup>er</sup> mai 2004. Ce sont Chypre, l'Estonie, la Hongrie, la Lettonie, la Lituanie, Malte, la Pologne, la République tchèque, la Slovaquie, la Slovénie.

## **1.2 Les coûts de production agricole pour l'analyse des politiques publiques**

L'analyse microéconomique des coûts de production agricoles, qu'elle soit rétrospective ou prospective, est un outil d'analyse des marges nettes, et donc permet d'évaluer la compétitivité des agriculteurs, un des éléments majeurs du développement ou du maintien de filières agro-alimentaires dans les régions européennes. L'estimation des coûts de production permet en effet d'apporter certains éléments de réponse aux questions que pose l'adaptation de l'agriculture française au contexte des marchés agricoles qu'ils soient nationaux, européens ou internationaux :

### **1.2.1 La libéralisation des échanges internationaux**

Les agriculteurs français, deviendront vraisemblablement de plus en plus dépendants des conditions régnant non seulement sur le marché européen, mais aussi sur les marchés mondiaux. En particulier, ils devront faire face à une variabilité accrue des prix agricoles. Une étude détaillée des coûts de production des différentes orientations technico-économiques des exploitations agricoles permet d'apprécier les possibilités d'adaptation de l'agriculture française à une concurrence de plus en plus vive, dans un contexte productif où les instruments de régulation s'élaborent désormais à un niveau supranational ;

## 1.2.2 Les réformes successives de la Pac

Dans un contexte de marché unique, la compétitivité relative des différentes filières productives (et donc le niveau relatif de leurs coûts de production) contribuera de plus en plus à la transformation du paysage rural des régions françaises par son impact sur l'évolution des structures en termes de spécialisation productive et de dimension économique des exploitations agricoles ;

## 1.2.3 L'élargissement de l'Union européenne

La comparaison des coûts de production entre les États membres met en évidence leurs avantages comparatifs respectifs et permet ainsi d'analyser l'évolution des équilibres de marché au niveau d'une Union élargie.

# 2. Méthodologies d'estimation des coûts de production agricole

## 2.1 La reconstitution des affectations selon les normes de comptabilité analytique

L'estimation des coûts de production repose à l'origine sur l'enregistrement direct de la comptabilité détaillée des exploitations agricoles, à partir de l'affectation a priori des charges aux différentes productions que réalisent les différentes méthodes de comptabilité analytique. Cependant, pour des questions de coût d'élaboration de ces comptabilités analytiques, ces estimations sont souvent réalisées sur un nombre restreint de produits concernant l'atelier de production ciblé. D'autre part, elles sont en général effectuées sur une base limitée à un réseau technico-économique de producteurs spécialisés, localisé dans une zone géographique déterminée.

D'autre part, bien qu'elles prétendent parfois au statut d'observations, ces estimations relèvent plutôt d'une reconstitution a posteriori de l'affectation des charges. Elles demeurent donc largement dépendantes des clefs de répartition utilisées fonction des normes techniques et financières qui président à l'établissement des comptabilités analytiques. Pour cette raison, on recense une pluralité de méthodes, se déclinant selon diverses combinaisons croisant système productif et adaptation du plan comptable agricole. Au plan européen ou international, malgré les efforts de normalisation des systèmes d'information comptables sur les entreprises, il n'existe pas de norme en la matière. Cependant, pour des produits comme les céréales, certaines méthodologies d'estimation, comme la méthode *TYPICAL* utilisée par l'*United States Department of Agriculture (USDA)*, servent de facto de référence malgré les difficultés d'adaptation à des contextes parfois très différents. Comme exemple de ce type de méthodologie, on peut citer l'enquête céréalière réalisée par l'Office national interprofessionnel des céréales (**Onic**) auprès d'un échantillon empirique d'agriculteurs spécialisés (enquête ECS) provenant de cinq départements du Bassin parisien [1].

En raison de la grande diversité des produits et des territoires, l'observation directe des coûts de production des exploitations agricoles supposerait, dans le contexte français, une mobilisation de moyens d'enquête disproportionnés par rapport à l'objectif poursuivi. De telles approches ne semblent donc adaptées ni à l'estimation de coûts de production pour des exploitations agricoles présentant de multiples combinaisons productives ni aux comparaisons inter-régionales dans un contexte européen.

## 2.2 Evaluation microéconomique des coûts de production à partir du Rica

### 2.2.1 La base empirique de l'estimation

Obtenir une estimation par produit des charges unitaires sur la base d'une enquête auprès d'un échantillon représentatif d'exploitations agricoles semble une procédure plus rationnelle pour fournir

une évaluation des coûts de production des principaux produits agricoles. Outil privilégié d'observation microéconomique des exploitations agricoles grâce aux informations comptables qu'il procure sur leur gestion, le Réseau d'information comptable agricole (**Rica**) est une enquête européenne annuelle [10], basée en France sur un échantillon d'environ 7 000 exploitations représentant 350 000 exploitations professionnelles [9]. Sa représentativité est assurée par un plan de sélection élaboré selon la méthode des quotas sur la base d'une stratification de la population des exploitations agricoles qui est ajustée tous les deux ans à partir d'estimations fournies par l'Enquête structure des exploitations agricoles (**Esea**). Cependant, le Rica ne propose pas de comptabilité analytique qui permettrait d'affecter directement les charges opérationnelles aux différentes productions pour en déduire les marges brutes des différentes spéculations agricoles. Les données comptables dont on dispose par le biais du Rica fournissent des informations sur les charges globales payées par l'exploitant agricole et les produits bruts générés par les différentes productions. Ces informations comptables sont ventilées selon la nature de la charge et de la production, mais les charges ne sont pas affectées aux différentes productions.

L'évaluation des coûts unitaires de production afférentes par unité de surface ou de cheptel pour chaque spéculation peut être effectuée sur la base d'une affectation des charges effectivement supportées par l'exploitation proportionnellement au produit dégagé par chaque production agricole présente sur l'exploitation. On obtient ainsi des coûts unitaires par unité monétaire de produit qu'il est possible ensuite de convertir en coûts unitaires par unités de poids, de surface ou d'effectif en rapportant le produit dégagé par le bien  $j$  aux quantités produites, respectivement aux surfaces cultivées ou à l'effectif de bétail.

En utilisant les estimations de ces coefficients techniques de production par unité de surface ou d'effectif, on peut déduire les charges spécifiques du produit et ainsi obtenir une estimation de la marge brute unitaire pour chaque spéculation. Si l'on déduit les charges de structure de la marge brute, on obtient alors la marge nette qui peut être assimilée à un revenu d'exploitation hors rémunération du travail.

**Tableau 1 : la matrice d'input-output des coefficients techniques de production**

CHARGES	PRODUITS	$X_1$	...	$X_j$	...	$X_J$	TOTAL
$Y_1$							$Y_1(i)$
$\vdots$							$\vdots$
$Y_k$				$a_k^j$			$Y_k(i)$
$\vdots$							$\vdots$
$Y_K$							$Y_K(i)$
TOTAL		$X_1(i)$	...	$X_j(i)$	...	$X_J(i)$	$\Sigma_X = \Sigma_Y$

## 2.2.2 Le principe de l'estimation des coefficients techniques de production

Il s'agit donc de fournir une estimation des coefficients techniques de production, élément de la ligne  $k$  et de la colonne  $j$  d'une matrice d'input-output croisant en ligne les charges  $Y_k$  avec en colonne les productions  $X_j$ , dont on ne connaît pour l'exploitation agricole  $i$  que les marges : le total-ligne  $Y_k(i)$ , donnant le montant total de la  $k^e$  charge ; le total-colonne  $X_j(i)$ , donnant le produit total de la  $j^e$  production.

Les techniques de régression multiple permettent d'obtenir une estimation statistique de ces coefficients techniques de production en régressant les charges  $Y_k$ , considérées comme variables endogènes ou expliquées sur les produits  $X_j$  qui jouent alors le rôle de variables exogènes ou explicatives.

## 2.2.3 Les concepts de produit, charges et marge nette

La spécification d'un modèle d'estimation des coûts de production suppose de préciser les concepts de produit, de charges et de marge nette à partir desquels sera jugée la pertinence du modèle :

- la valeur  $X_j$  de la production d'un bien  $j$  est définie par le produit brut, somme algébrique des ventes, des variations de stock et de l'autoconsommation ;
- les charges  $Y_k$  portent sur des facteurs de production  $k$  et regroupent les consommations intermédiaires, les frais d'exploitation et les amortissements ;
- la marge nette  $M$  de l'exploitation agricole  $i$  est défini par la somme pour l'ensemble des  $p$  spéculations du différentiel entre le prix unitaire  $P_j$  d'un bien  $j$  et son coût unitaire  $C_j$  multiplié par le volume  $Q_j$  de sa production.

$$(1) \quad M = \sum_{j=1}^p (P_j - C_j) \times Q_j$$

On distingue deux types de charges : d'une part, les charges spécifiques portant sur des facteurs dont le niveau d'intrant est directement lié aux volumes de production de l'exploitation agricole (semences, engrais, aliments du bétail, frais vétérinaires, etc.) ; d'autre part, les charges de structure qui dépendent de la taille de l'exploitation et de l'outil de production (coût foncier, frais financiers, entretien du matériel et des bâtiments, etc.).

## 3. L'estimation économétrique des coûts de production

S'inspirant d'un modèle économétrique visant à reconstituer une matrice d'input-output [13], les travaux menés dès 1983 à la division Agriculture de l'Insee fournissent une évaluation microéconomique des marges brutes unitaires par l'intermédiaire d'une estimation statistique des coûts de production. Depuis cette date, différentes spécifications de ce modèle économétrique ont été utilisées dans des contextes distincts.

### 3.1 Spécification initiale du modèle d'estimation des coûts de production agricoles à partir du Rica français

La spécification du modèle est basée sur une fonction de coût linéaire, de type Léontieff, à coefficients constants : les charges sont proportionnelles aux produits bruts afférents. Pour chaque charge  $Y_k$ , la version initiale du *modèle d'estimation des coûts de production agricoles (Mecpa-I)* cherche à estimer

sur chacun des produits bruts  $X_j$ , les coefficients techniques  $a_k^j$  dans l'équation de régression suivante appliquée à l'ensemble des exploitations du Rica :

$$(2) \quad Y_k = \sum_{j=1}^p a_k^j X_j + e_k \text{ avec } e_k \text{ i.i.d.}$$

La version initiale utilise la méthode d'estimation des *moindres carrés ordinaires (MCO)*. Cette méthode numériquement adaptée à l'estimation des fonctions de coûts linéaires, fournit l'estimateur sans biais de variance minimale des coefficients techniques de production, pourvu que les hypothèses du modèle soient respectées (erreurs aléatoires  $e_{k(i)}$  indépendantes et identiquement distribuées selon une loi normale  $N(0, \mathbf{s})$  de variance constante  $\mathbf{s}^2$ ).

### 3.2 Interprétation économique des hypothèses de spécification du modèle

Cette spécification de type Léontieff de la fonction de coût s'appuie sur trois hypothèses fondamentales :

- L'utilisation du facteur de production  $k$  ne dépend, à un facteur près  $e_k$  aléatoire, que du produit.

$$(3) \quad Y_k = f(X_j) + e_k$$

Une même technique de production s'impose à toutes les exploitations agricoles. Le modèle ne prend donc pas en compte la spécificité des exploitations liée aux effets d'échelle ou à l'adoption d'une technologie particulière de production ;

- La valeur de l'input utilisé en un facteur  $k$  pour chaque produit  $j$  est proportionnelle à la valeur de la production.

$$(4) \quad f_k(X_j) = a_k^j X_j$$

L'hypothèse de proportionnalité suppose la linéarité de la fonction de coût. La valeur des inputs est proportionnelle à la valeur de la production : les rendements sont donc supposés constants. Le coût total observé est la somme des coûts relatifs pour les différents produits. Les coefficients sont constants sur l'échantillon, la technologie de production n'est supposée dépendre que du produit agricole non des caractéristiques de l'exploitation ;

- L'absence de contrainte d'approvisionnement pour les facteurs de production.

$$(5) \quad Y_k = \sum_{j=1}^p a_k^j X_j + e_k \text{ avec } e_k \text{ i.i.d.}$$

Le choix d'un modèle à erreurs aléatoires  $e_k$  indépendantes et identiquement distribuées (**i.i.d.**) implique que la consommation d'un facteur de production par une exploitation  $i$ , n'affecte pas la consommation de ce facteur par une autre exploitation  $i'$ .

En outre, la technique d'estimation retenue effectue la régression équation par équation. Il n'y a pas de relation inter-équations liant les coefficients estimés par le modèle. Cette technique ne permet donc pas d'introduire des contraintes comptables portant sur l'ensemble de l'exploitation.

L'étude initiale [2] est réalisée en appliquant ce modèle économétrique à l'échantillon du Rica français de 1970 à 1978 et porte sur quatorze produits distinguant :

- huit produits végétaux : céréales, maïs, pommes de terre, légumes frais, fruits, fleurs, plantes industrielles et vins ;
- six produits animaux : porcs, volailles, œufs, lait, bovins et autres herbivores.

Malgré les limitations intrinsèques à cette spécification, les estimations réalisées par Mecpa-I, comparées à d'autres sources, sont jugées globalement satisfaisantes, à l'exception des résultats portant sur la pomme de terre et les œufs. Les engrais pour la production végétale et les aliments pour la production animale y apparaissent comme les principaux postes de charges. L'introduction d'un modèle à équations simultanées est déjà envisagée comme perspective d'amélioration du modèle pour tenir compte des contraintes comptables portant sur l'ensemble de l'exploitation.

### 3.3 Première application du modèle au cadre européen

En 1985, dans un contexte de remise en cause du système des prix garantis formant le cœur du dispositif originel de la Pac, la Direction générale de l'agriculture (DG VI) de la Commission européenne, préoccupée par l'impact des différentiels de coûts de production sur les filières agro-alimentaires au sein de l'UE, commande une étude comparative étendue à l'ensemble du Rica européen sur la période 1979-1984.

En réponse à l'appel d'offre de la Commission, une équipe mixte Inra-Insee se charge de la première application du Mecpa au Rica européen ([4], [5] et [6]). L'étude porte sur onze produits agrégés : blé tendre, orge, maïs, oléagineux, betterave sucrière, légumes, fruits, bovins viande, bovins lait, porcins, volailles. Par rapport à la version initiale, la première version européenne du modèle, *Mecpa-EI*, comporte les spécifications supplémentaires suivantes :

- l'estimation est toujours effectuée équation par équation (i.e. charge par charge) au moyen des MCO, mais les résultats sont ensuite ajustés pour respecter, produit par produit, la contrainte comptable d'égalité entre le produit brut d'une part, et la somme des coûts et des marges d'autre part ;
- les coefficients non pertinents (par exemple, le coefficient correspondant aux aliments du bétail pour les produits végétaux) sont contraints a priori à une valeur nulle ;
- les coefficients négatifs sont contraints à une valeur nulle a posteriori si leur valeur absolue est négligeable ou s'ils reflètent une forte colinéarité entre variables explicatives, malgré le risque de biaiser les autres estimations.

Les principaux résultats de ce premier travail à l'échelle européenne sont les suivants :

- la comparaison [3] des coûts de production dans chaque État membre en parité de pouvoir d'achat<sup>1</sup> met en évidence, pour la période 1979-1984 trois niveaux de performance : d'une part la France, l'Irlande, les Pays-Bas et la Belgique présentent un coût global<sup>2</sup> inférieur de 5 à 10% à la moyenne communautaire ; d'autre part la Grèce, le Danemark, la RFA et l'Italie se situent au niveau moyen ; quant au Royaume-Uni, son indice de coût global est supérieur de 15 % à la moyenne ;
- l'analyse des coûts par produit permet de distinguer deux groupes de pays ; dans le premier groupe (Royaume-Uni, Danemark et Allemagne), les coûts de production sont proches de l'indice du coût global ; dans les autres pays, les avantages comparatifs sont inégalement répartis entre les orientations. Par exemple, le handicap de l'Italie en grandes cultures contraste avec sa compétitivité en fruits et légumes. Quant aux Pays Bas et à la Belgique, ils sont avantagés dans le domaine de la production animale. Cette disparité des coûts de production au niveau européen peut s'expliquer par l'hétérogénéité des conditions pédo-climatiques, par la diversité des systèmes de production et de la structure des exploitations, mais aussi par des particularismes nationaux dans leur gestion (recours plus important au travail salarié au Royaume-Uni, frais financiers plus importants au Danemark, amortissements plus élevés en Allemagne).

<sup>1</sup> Les résultats monétaires sont convertis en parité de pouvoir d'achat (PPA) à l'aide des coefficients standards de pouvoir d'achat (SPA) publié par Eurostat, l'Office statistique des communautés européennes.

<sup>2</sup> L'indice global des coûts estimés est calculé, pour un pays donné, en sommant les coûts de l'ensemble des produits de ce pays, pondérés par la part en valeur de chaque produit dans la production nationale. Le coût communautaire d'un produit est obtenu en sommant les coûts nationaux, pondérés par la part de chaque État dans la valeur de la production communautaire.



- la prise en compte des coûts fonciers et du travail familial dans le modèle n'apparaît pas satisfaisante, ce qui limite les possibilités de comparaison des coefficients techniques estimés par Mecpa-E1 entre les différents produits.

### 3.4 Introduction des contraintes comptables a posteriori dans l'estimation MCO

Fruit de la collaboration entre la DG VI, l'Inra et l'Insee, une nouvelle étape dans le développement du modèle est franchie avec la publication en 1990 d'une deuxième étude ([7] et [8]) complétant et actualisant les résultats de la précédente :

- Les estimations du modèle sont toujours obtenues par la méthode des MCO appliquée équation par équation. Mais les résultats sont désormais redressés a posteriori pour respecter des contraintes inter-équations issues de l'application des règles comptables à chaque exploitation ;
- la marge nette s'évalue à un facteur aléatoire près  $v$  comme la somme, sur l'ensemble des  $p$  spéculations, des revenus générés par chaque produit.

$$(6) \quad M = \sum_{j=1}^p \mathbf{b}^j X_j + \mathbf{e} \quad \text{avec } \mathbf{e} \text{ iid}$$

- la définition de la marge nette comme différentiel entre la valorisation et le coût du produit, impose une contrainte linéaire sur les coefficients pour tout bien  $j$ . Il s'agit ici de respecter la logique comptable.

$$(7) \quad \sum_{k=1}^K a_k^j + b^j = 1 \quad \forall j$$

Les estimations portent dorénavant sur la période 1984-1986. L'application de ce deuxième modèle européen, *Mecpa-E2*, est élargie aux niveaux sectoriel (notamment pour les fruits et la viticulture) et géographique (pour intégrer l'Espagne et le Portugal, nouveaux États membres de la CEE). La mise en œuvre du modèle autorise désormais, pour les principaux produits<sup>3</sup>, une analyse des coûts par grandes régions agricoles et non plus seulement par pays. Le calcul des coûts liés au travail familial est amélioré.

### 3.5 Les problèmes rencontrés dans l'estimation des MCO redressée a posteriori

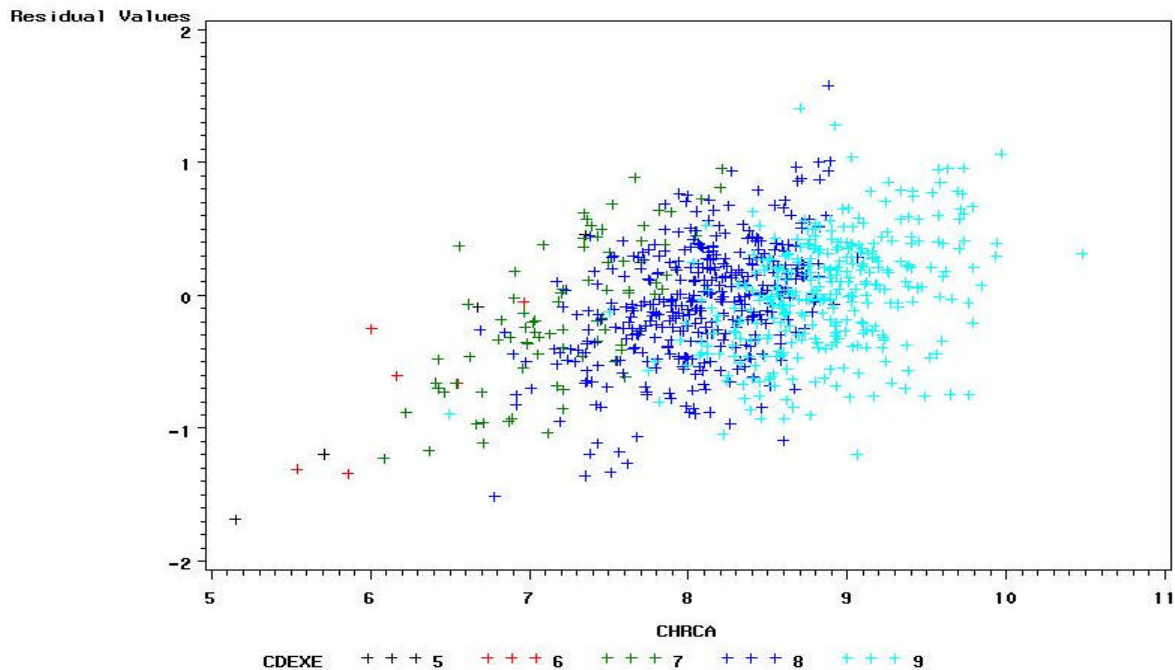
Les principaux problèmes d'estimation constatés sont une forte variance des estimateurs, des coefficients de régression négatifs (alors qu'ils devraient être positifs du fait de leur signification économique), des termes constants significativement différents de zéro (peu vraisemblables dans le cas de charges spécifiques) et des coefficients sous-estimés ou instables dans le temps. Ces problèmes de précision des estimateurs trouvent leur source dans le non respect des hypothèses du modèle :

- l'hétéroscédasticité se manifeste pour certaines équations de charge par une dispersion des résidus variant en fonction du niveau de production repéré par la classe de dimension économique (cf. figure 2) ; pour l'équation sur les charges en carburants (CHRCA), on constate que la dispersion des résidus varie en fonction du niveau de l'input et de la dimension économique de l'exploitation que représente la CDEXE, classe de dimension économique. On a donc,  $V(\mathbf{e}_{kji}) \neq \mathbf{s}^2$ , violation de l'hypothèse de variance constante et de celle d'indépendance des erreurs aléatoires,

<sup>3</sup> Blé tendre, blé dur, orge, maïs grain, betterave sucrière, pomme de terre, viticulture, lait et viande bovine, ovins.

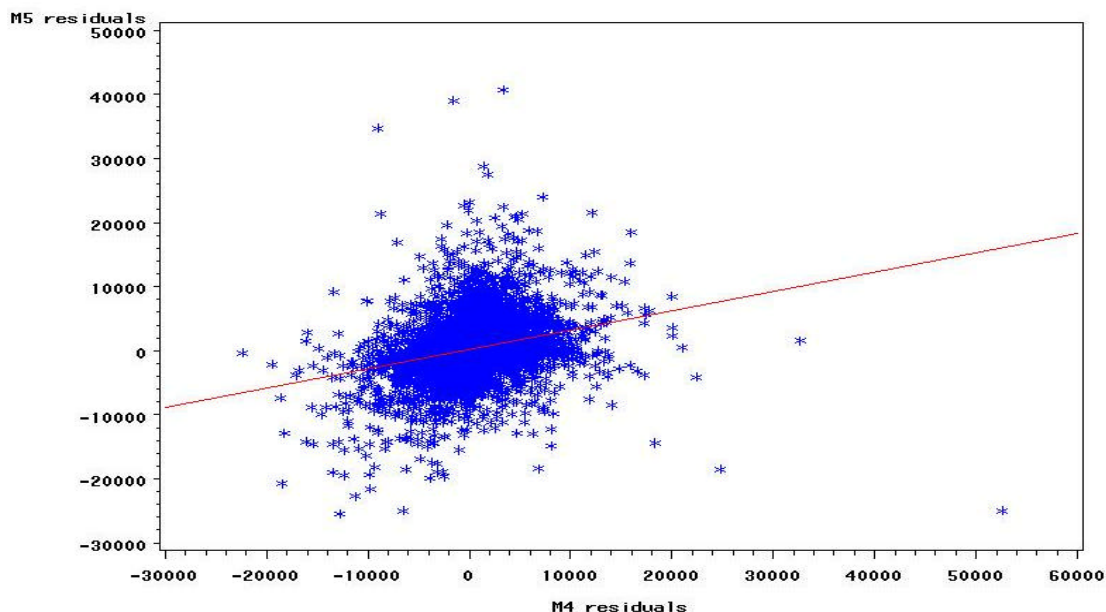
$E(\mathbf{e}_{ki}, \mathbf{e}_{k'i'}) \neq 0$ . La matrice de variance-covariance des erreurs n'est donc ni diagonale ni constante pour certaines équations de charge,  $Cov(\mathbf{e}_k) \neq \mathbf{S}^2 I$ . En effet, l'hypothèse d'une matrice de variance-covariance diagonale n'est pas réaliste en raison de l'existence d'effets climatiques ou pédologiques pouvant affecter plusieurs individus d'une même région.

**Figure 2 : valeurs résiduelles pour l'équation prédictive de la charges en carburants en fonction de la variable endogène (CHRCA), selon les classes de dimension économique croissantes (CDEXE)**



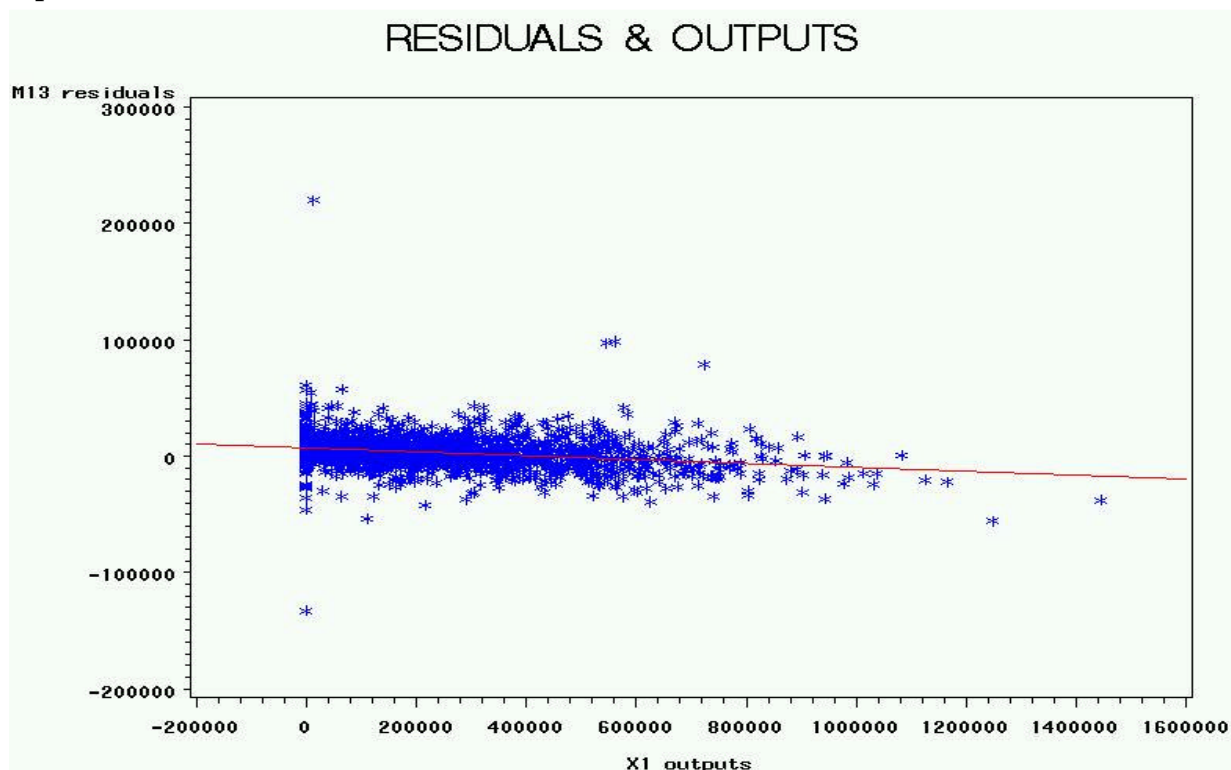
- La corrélation entre les résidus des différentes équations de charge du modèle ( $Cov(\mathbf{e}_k, \mathbf{e}_{k'}) \neq 0$ ) induit des biais dans l'estimation des coefficients ; à l'origine de ces corrélations, le fait que le montant de la charge  $Y_k$  peut ne pas être uniquement déterminé par les produits bruts  $X_j$  des biens afférents. En effet, d'autres facteurs peuvent influencer sur le niveau des intrants. L'environnement de l'exploitation (nature du sol, insertion économique) ou certaines caractéristiques liées à l'exploitation (information, stratégie productive) vont affecter simultanément le niveau de plusieurs charges (cf. l'exemple des engrais et des phytosanitaires en figure 3). Lorsqu'elle est appliquée, la contrainte comptable renforce ces dépendances entre les résidus des équations de charges en imposant que, pour un même intrant, la somme des coefficients techniques soit égale à l'unité.

**Figure 3 : corrélation ( $r=0,275$ ) entre les résidus des équations de charges d'engrais (M4) et de phytosanitaires (M5).**



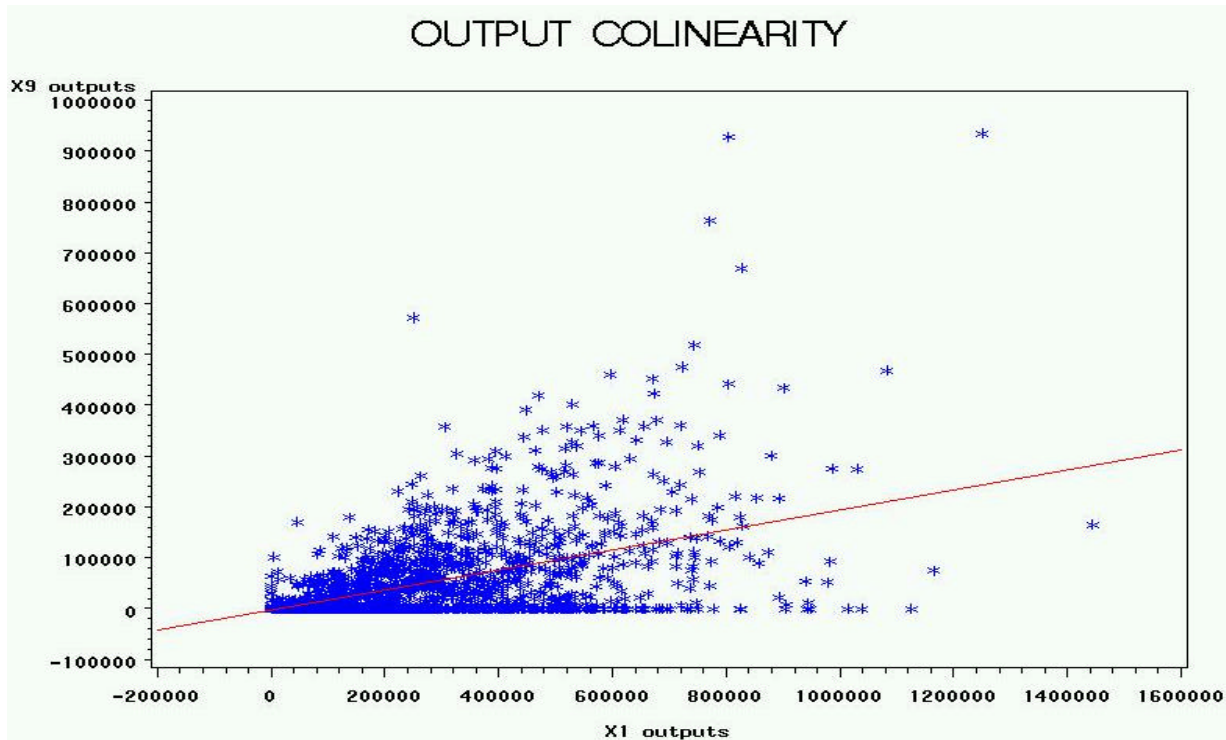
- l'hypothèse d'orthogonalité entre les régresseurs  $X_j$  et les perturbations aléatoires (k n'est pas toujours vérifiée dans l'estimation de certaines charges comme entre le blé tendre et les résidus de l'équation des charges d'assurance (cf. figure 4). En effet les quantités produites sont déterminées en liaison avec le niveau d'intrant choisi par l'exploitant. Les estimateurs des coefficients peuvent donc de ce fait être biaisés. L'utilisation des méthodes à variables instrumentales comme solution à ce problème est fortement limitée par la difficulté d'obtenir des variables instrumentales adéquates. Divay et Meunier [15] suggèrent d'utiliser les salaires et les taxes mais, en agriculture, ces variables instrumentales sont faiblement corrélées aux productions.

**Figure 4 : corrélation ( $r = - 0,25$ ) entre les résidus des équations de charges d'assurance (M13) et le produit brut du blé tendre (X1).**



- Les écart-types importants des coefficients estimés traduisent la faible qualité de certaines des estimations réalisées. La faible vraisemblance des valeurs obtenues peut s'expliquer par des phénomènes de multicollinéarité. La multicollinéarité des régresseurs provient des associations productives (rotations culturales standards) pratiquées par la grande majorité des producteurs, en particulier dans l'orientation grandes cultures, comme ici, entre le blé tendre et le colza (cf. figure 5). Ce phénomène de multicollinéarité produit des coefficients peu fiables, notamment pour le blé tendre, l'orge, la pomme de terre, la betterave et les oléagineux. Il se renforce lors d'estimations faites sur les orientations technico-économiques ou les régions spécialisées. La multicollinéarité permet également d'expliquer les coefficients négatifs obtenus.

**Figure 5 : corrélation (  $r = 0,50$ ) entre le produit brut du blé tendre (X1) et le produit brut du colza (X9).**



La détection des situations de multicollinéarité s'effectue au moyen des indicateurs suivants :

- le facteur d'inflation de la variance ou *VIF* (*Variance Inflation Factor*) où  $R_j^2$  est le coefficient de corrélation multiple de la régression avec constante de la variable exogène  $X_j$  sur les autres régresseurs ;

$$VIF = \frac{1}{1 - R_j^2}$$

- l'indice de conditionnement maximal proposé par Besley, Kuh et Welsch utilisant le rapport des valeurs propres extrêmes de la matrice de corrélation des variables exogènes

$$m_j = \sqrt{\frac{I_{\max}}{I_{\min}}}$$

En vue d'améliorer la qualité des résultats de ce second modèle européen, Mecpa-E2, J-C. Bureau (cf. [7]) teste empiriquement d'autres méthodes d'estimation des coefficients de régression (moindres carrés généralisés, modèles à équations simultanées, doubles moindres carrés,...). L'étude comparée des techniques d'estimation le conduit à préconiser l'utilisation de la procédure d'estimation des moindres carrés généralisés (MCG) pour tenir compte de l'hétéroscédasticité, et des estimateurs bornés (« ridge ») en raison des problèmes induits par la multicollinéarité des régresseurs.

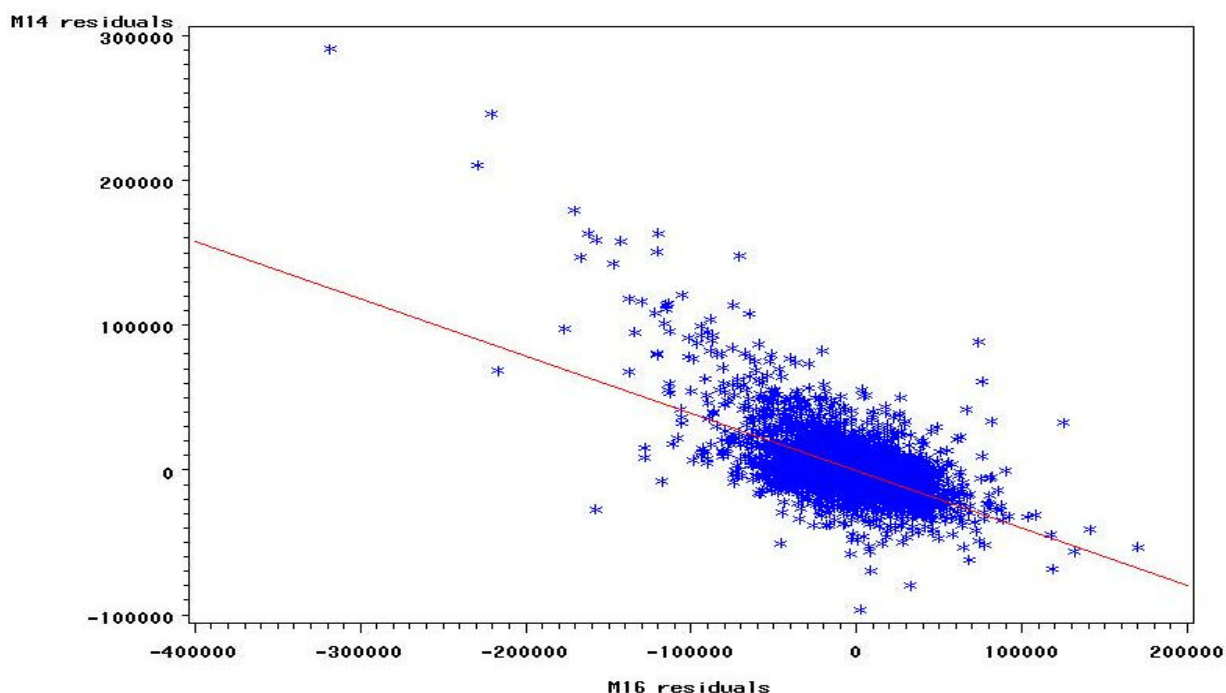
### 3.6 Développement du modèle à équation simultanée et applications

En Europe, l'évaluation de la réforme de la Pac intervenue en 1992-93 et la promulgation du *Farm Bill* en 1996 aux *USA* ravivent l'intérêt des pouvoirs publics pour l'analyse comparée des coûts de production. Leur estimation sur une base européenne permet d'effectuer des simulations de revenu pour analyser l'impact des différents scénarios de réforme de la Pac.

Ses spécifications sont donc revues pour prendre en compte le nouveau mode d'attribution des subventions institué à l'occasion de cette réforme. Une collaboration tripartite Inra-Insee-SCEES est engagée afin d'améliorer la procédure d'estimation des coefficients techniques décrivant l'affectation des charges aux produits [17]. Fin 1995, le SCEES est chargé de la maintenance du Mecpa afin d'internaliser son utilisation au ministère de l'Agriculture et de la Pêche. Les principales améliorations apportées au modèle sont les suivantes :

- pour tenir compte des fortes corrélations inter-équations (cf. figure 6) et pour intégrer dans l'estimation des coûts les contraintes globales de l'exploitation, la technique de régression équation par équation est remplacée par une procédure d'estimation d'un système à équations simultanées, de type *SUR* (*Seemingly Unrelated Regression*) [26];

**Figure 6: corrélation ( $r = -0,61$ ) entre les résidus des équations de marge nette (M16) et de charges (M14).**



- pour pouvoir calculer des dispersions sur les coûts par produit ou des coûts moyens sur des sous-échantillons<sup>4</sup>, un module «d'imputation des coûts individuels» a été ajouté au modèle ; pour chaque exploitation et chaque poste de charge, ce module affecte les résidus proportionnellement aux différents produits afin d'ajuster la charge estimée à la charge observée (pour une justification détaillée de cette affectation proportionnelle, cf. [18] et [14]).

<sup>4</sup> Par exemple, selon la région, l'orientation technico-économique (Otex) ou la classe de dimension économique (Cdex).

La mise en œuvre de ces améliorations aboutit à la version standard du modèle d'estimation des coûts de production agricole (**Mecpa-ST**) actuellement en service. La spécification du modèle Mecpa-ST a été présentée initialement en interne dans [18], puis publiée de façon détaillée dans [14]. Elle a été successivement utilisée pour étudier :

- la formation et la répartition des gains de productivité dans l'agriculture française par la méthode des comptes de surplus [9] ;
- l'évolution et la dispersion des coûts des exploitations françaises en grandes cultures sur la période 1981-1994 [17] ;
- les coûts de production spécifiques à certains produits comme le blé tendre, la pomme de terre ou le porc [12] ;
- la compétitivité de l'agriculture française, sept ans après la réforme de la Pac de 1992, et ses stratégies d'adaptation face aux accords de Berlin de mars 1999 [19] ;
- l'évolution sur la période 1991-2001 des coûts de production de six produits agricoles majeurs représentant les principales organisations communes de marché existant dans l'espace européen [13] ;
- l'impact sur les revenus agricoles des différents scénarios en discussion dans le cadre de la révision de la PAC à mi-parcours et du cycle de négociations de l'OMC initié à Doha ;
- l'évolution comparée des coûts de production français, néerlandais et hongrois, pour évaluer en particulier les conséquences de l'élargissement de l'UE sur les prix, coûts et marges des agriculteurs hongrois<sup>5</sup>.

En outre, l'opération de révision décennale des marges brutes standards, indicateur normalisé au plan européen du potentiel productif des exploitations agricoles, prend désormais comme point de départ les coûts de production régionaux estimés par le modèle selon une nomenclature de charges et de produits aussi proche que possible des spécifications de l'acquis communautaire en matière de statistique agricole ( cf. [20] et [14]).

## **4. La modélisation selon les moindres carrés partiels**

Historiquement, le modèle de la régression PLS trouve son origine dans l'approche « soft modeling » proposée par Herman Wold dans le cadre de l'étude des modèles de relations structurelles sur variables latentes (*path model with latent variables*). Dans cette approche de « modélisation douce », on construit des modèles en utilisant des techniques d'ajustement numérique du type moindres carrés sans faire appel à des hypothèses probabilistes. Aux fondements de cette approche, se situe un algorithme itératif des moindres carrés proposé par Wold en 1966 [25], *NILES - Non linear Estimation by Iterative Least Squares*, permettant d'effectuer une analyse en composantes principales d'un tableau comportant des données manquantes.

### **4.1 La régression des moindres carrés partiels comme modélisation sur variables latentes**

Souvent dans les situations de multicolinéarité ou de saturation du modèle, bien que les variables explicatives a priori soient nombreuses, seul un petit nombre de facteurs structurels, généralement inobservables, permet de rendre compte de la variabilité des variables expliquées. L'approche de la « modélisation douce » consiste à extraire ces facteurs appelés variables latentes, par combinaison

---

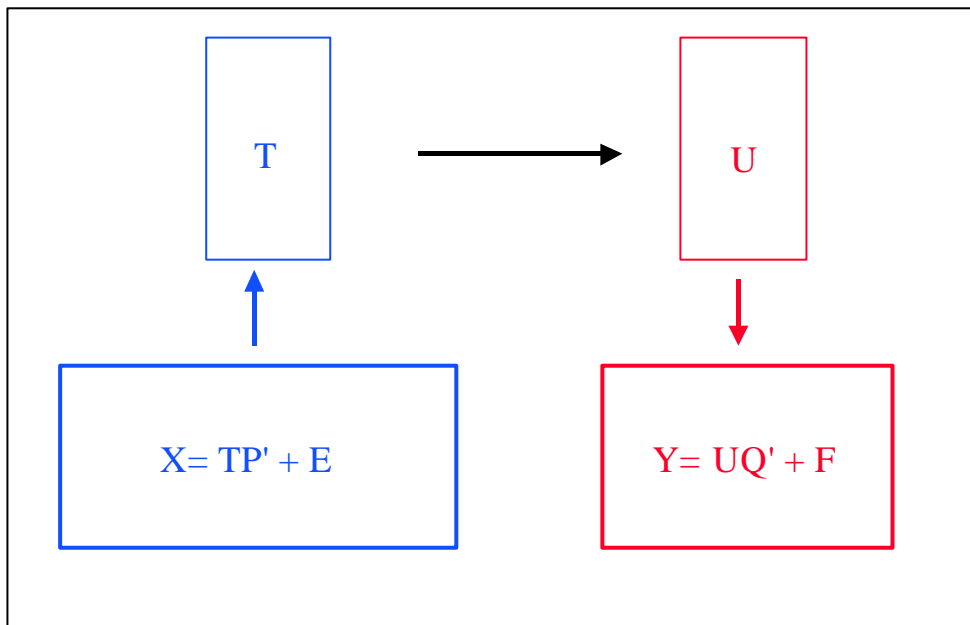
<sup>5</sup> Travaux menés par le bureau de l'Analyse économique et de la Prospective (ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires rurales, direction des Affaires financières) dans le cadre d'un projet PHARE (n°HU9806-03-02) pour la reprise de l'acquis communautaire en matière de système d'information agricole.

linéaire des variables exogènes, en sélectionnant ceux qui modélisent au mieux le comportement des variables endogènes. Notre problème est d'expliquer  $Y$ , tableau des variables endogènes, par  $X$  tableau des variables exogènes en présence de multicollinéarité. La solution est de passer par des composantes orthogonales de  $X$  pour réaliser la projection de  $Y$  sur  $X$ . Une illustration immédiate de cette approche est la régression de  $Y$  sur les composantes principales de  $X$ . Cette méthode comporte trois étapes :

1. calcul des composantes orthogonales de  $X$  ;
2. régression de  $Y$  sur les composantes orthogonales de  $X$  les mieux corrélées à  $Y$  ;
3. ré-écriture de l'équation de régression en fonction de  $X$ .

Cependant les composantes principales de  $X$ , orientées vers la description du tableau  $X$ , ne sont pas forcément les meilleures variables latentes pour expliquer la variabilité contenue dans le tableau  $Y$ . Afin de pallier à cet inconvénient, la modélisation sur variables latentes de la régression des moindres carrés partiels (**MCP**) ou régression **PLS** (*Partial Least Squares*) [21] procède de manière indirecte en extrayant des pseudocomposantes  $T$  à partir des variables exogènes  $X_j$  pour estimer des pseudocomposantes  $U$  qui permettront de calculer les variables endogènes  $Y_k$ . La solution de la régression **PLS** est de passer par des composantes orthogonales  $T$  dans  $X$  et  $U$  dans  $Y$  en maximisant leur covariance sous contrainte d'orthonormalisation.

**Figure 7 :** relation entre composantes et variables



- $X$  : matrice des variables exogènes
- $T$  : matrice des  $X$ -pseudocomposantes
- $P$  : matrice des  $X$ -saturations
- $E$  : matrice des  $X$ -résidus
- $Y$  : matrice des variables endogènes
- $U$  : matrice des  $Y$ -pseudocomposantes
- $Q$  : matrice des  $Y$ -saturations
- $F$  : matrice des  $Y$ -résidus

Dans la régression des moindres carrés partiels, les composantes  $T$  et  $U$  sont choisies pour obtenir une liaison optimale au sein de chaque paire de composantes. Il s'agit en fait d'une forme robuste de régression sur variables latentes qui recherche les directions de l'espace des variables exogènes liées aux plus fortes variations dans l'espace des variables endogènes mais biaisées vers des directions mieux prédites que dans la régression sur les composantes principales. La régression selon les moindres carrés partiels s'appuie sur la décomposition spectrale de l'opérateur  $X'Y$  où  $X$  est le tableau des variables exogènes et  $Y$ , le tableau des variables endogènes, proposant ainsi une reconstitution hiérarchisée de la variabilité du tableau  $X$  orientée vers la prévision du tableau  $Y$  grâce au critère de maximisation de la covariance des pseudo-composantes  $T$  et  $U$  sous contrainte de normalisation.

Si le nombre de composantes extraites est supérieur ou égal à la dimension de l'espace des facteurs, la régression MCP fournit alors la solution des MCO. Cependant, la méthode des moindres carrés partiels permet en général de se limiter à un nombre restreint de composantes. Le choix du nombre de composantes à extraire est effectué sur la base de techniques heuristiques cherchant à minimiser la variance résiduelle. Une des approches utilisées consiste à construire le modèle avec un nombre fixé de variables explicatives pour un sous-ensemble particulier des données, puis à tester le modèle obtenu sur un autre sous-ensemble des données en choisissant le nombre de composantes qui minimise l'erreur d'estimation. On peut également choisir le plus petit nombre de composantes associé aux modèles dont la variance résiduelle n'est pas significativement supérieure à celle du modèle à erreur résiduelle minimale [24]. Si l'on ne dispose pas d'un échantillon-test, on peut utiliser chaque observation en tant qu'échantillon-test, on effectue alors une *validation-croisée*.

Le critère empirique utilisé pour la validation du nombre  $L$  de composantes d'un modèle est le  $PRESS^6$  ou somme des carrés des erreurs d'estimation, soit en validation-croisée :

$$PRESS = \sum_{i=1}^L (y_i - \hat{y}_{\{-i\}})^2$$

La règle est de retenir la  $l^e$  composante si on obtient une diminution du  $PRESS$  d'un pourcentage jugé significatif, par exemple 5% :

$$\sqrt{PRESS_l} \leq 0,95 * \sqrt{PRESS_{l-1}} .$$

**Tableau 2 : Validation du nombre de composantes pour le modèle des coûts de production selon la procédure PLS de SAS.**

The PLS Procedure: Random Subset Validation for the Number of Extracted Factors			
Number of Extracted Factors	Root Mean PRESS	Prob > PRESS	
0	1.155714	<.0001	
1	1.067863	<.0001	
2	1.008234	<.0001	
3	0.938102	<.0001	
4	0.920035	<.0001	
5	0.919288	<.0001	
6	0.923448	<.0001	
7	0.903448	<.0001	
8	0.890096	<.0001	
9	0.889128	<.0001	
10	0.885717	<.0001	
11	0.891929	<.0001	
12	0.872239	0.0550	
13	0.876255	0.4340	
14	0.87337	0.2300	
15	0.871804	1.0000	
Minimum root mean PRESS			0.8718
Minimum number of factors			15
Smallest number of factors with p > 0.1			13

En suivant cette règle empirique, nous sélectionnons seulement les quatre premières composantes. En utilisant le test de van der Voet [24], on sélectionne les treize premières composantes.

<sup>6</sup> Predicted Residual Sum of Squares



## 4.2 Les algorithmes de la régression des moindres carrés partiels

Il existe plusieurs algorithmes permettant d'obtenir une solution des moindres carrés partiels, chacun ayant son objectif et son cadre d'application propre :

- *NIPALS 1* : algorithme itératif d'analyse en composantes principales sans données manquantes ;
- *NIPALS 2* : estimation d'un modèle bilinéaire avec données manquantes ;
- *PLS 1* : régression MCP pour  $y$  endogène univarié ;
- *PLS 2* : régression MCP pour  $Y$  endogène multivarié ;
- *SIMPLS* : algorithme de résolution rapide.

Dans l'application des coûts de production où l'on souhaite résoudre le système à équations simultanées constitué de la matrice d'input-output, l'ensemble  $Y$  des charges est évidemment multivarié et nous appliquerons donc la méthode *PLS2*. La procédure *PLS* de SAS [23] ne proposant qu'une solution approchée basée sur l'algorithme *SIMPLS*, nous utilisons un macroprogramme *PLS2* dont l'algorithme itératif est résumé dans le tableau 2 et complètement décrit en annexe 1. Il s'inspire des travaux de Michel Tenenhaus ([21]). L'ensemble des tableaux ne présentant pas de données manquantes, nous n'utiliserons pas l'algorithme *NIPALS 2*, qui conserve néanmoins son intérêt dans une telle éventualité.

L'algorithme *PLS2* consiste à appliquer alternativement et itérativement une série de régressions successives à un vecteur initial pour extraire les pseudo-composantes respectives de chaque tableau jusqu'à convergence du processus.

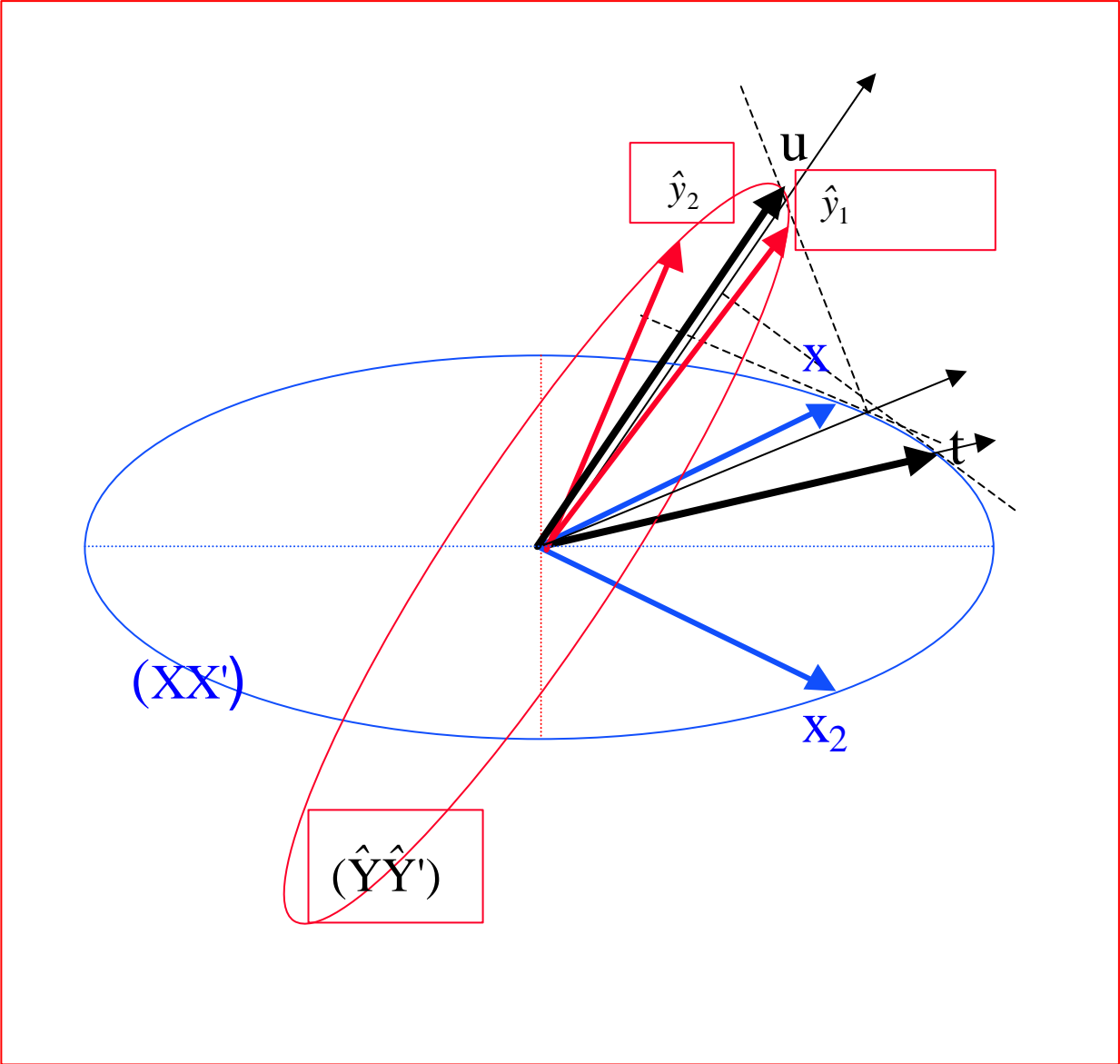
Lorsqu'on a des données manquantes, on applique le principe de l'algorithme *NIPALS* à l'algorithme *PLS2* : les coordonnées des vecteurs  $w_k$ ,  $t_k$ ,  $c_k$  et  $u_k$  sont calculées comme les pentes des droites des moindres carrés passant par l'origine sur les données disponibles.

L'algorithme de régression *PLS1* est un cas particulier de l'algorithme de régression *PLS2* lorsque le tableau  $Y$  est univarié.

**Tableau 3 : Synthèse de l'algorithme PLS2 (d'après [21])**

<b>1</b>	"Initialisation": $X_0 = X$ ; $Y_0 = Y$ ;	
<b>2</b>	Pour $k=1, 2, \dots, K$ faire { début <b>2.1</b> $u_k = Y_{k-1} [ \cdot ; 1 ]$ ;	
<b>2.2</b>	Faire jusqu'à $ w_k - w_k^*  < \epsilon$ ? début	
<b>2.2.1</b>	$w_k = X'_{k-1} u_k / u'_k u_k$ ;	* projection de $u$ sur $X$
<b>2.2.2</b>	$w_k = w_k /  w_k $ ;	* normalisation de $w$
<b>2.2.3</b>	$t_k = X_{k-1} w_k / w'_k w_k$ ;	* calcul de $t$
<b>2.2.4</b>	$c_k = Y'_{k-1} t_k / t'_k t_k$ ;	* projection de $t$ sur $Y$
<b>2.2.5</b>	$u_k = Y_{k-1} c_k / c'_k c_k$ ; fin }	* calcul de $u$
<b>2.3</b>	$p_k = X'_{k-1} t_k / t'_k t_k$ ;	* projection de $t$ sur $X$
<b>2.4</b>	$X_k = X_{k-1} - t_k p'_k$ ;	* mises à jour de $X$ et de $Y$
<b>2.5</b>	$Y_k = Y_{k-1} - t_k c'_k$ ; fin }	

Figure 8 : Interprétation géométrique du fonctionnement itératif de l'algorithme PLS2 (d'après [16]).



### 4.3 La régression des moindres carrés partiels des charges sur les produits

Outil de modélisation, la régression *PLS* constitue également un outil exploratoire des données. Ainsi, les graphiques pseudo-factoriels permettent une visualisation des données dans l'espace des pseudocomposantes.

Sur le graphique des deux premières *X*-pseudocomposantes  $t_1$  et  $t_2$  (figure 9), sont projetées les variables exogènes (tableau *X*) mais aussi les variables endogènes (tableau *Y*). Cela permet d'avoir un aperçu des inter-relations entre variables exogènes *X* et variables endogènes *Y* dans un sous-espace privilégiant la variabilité des variables exogènes *X*. On remarque que la première *X*-pseudocomposante oppose les productions animales ( $t_1 < 0$ ) aux productions végétales ( $t_1 > 0$ ). La seconde composante  $t_2$  semble devoir différencier les produits selon l'importance de la marge nette dégagée (variable *Y16* Revenus).

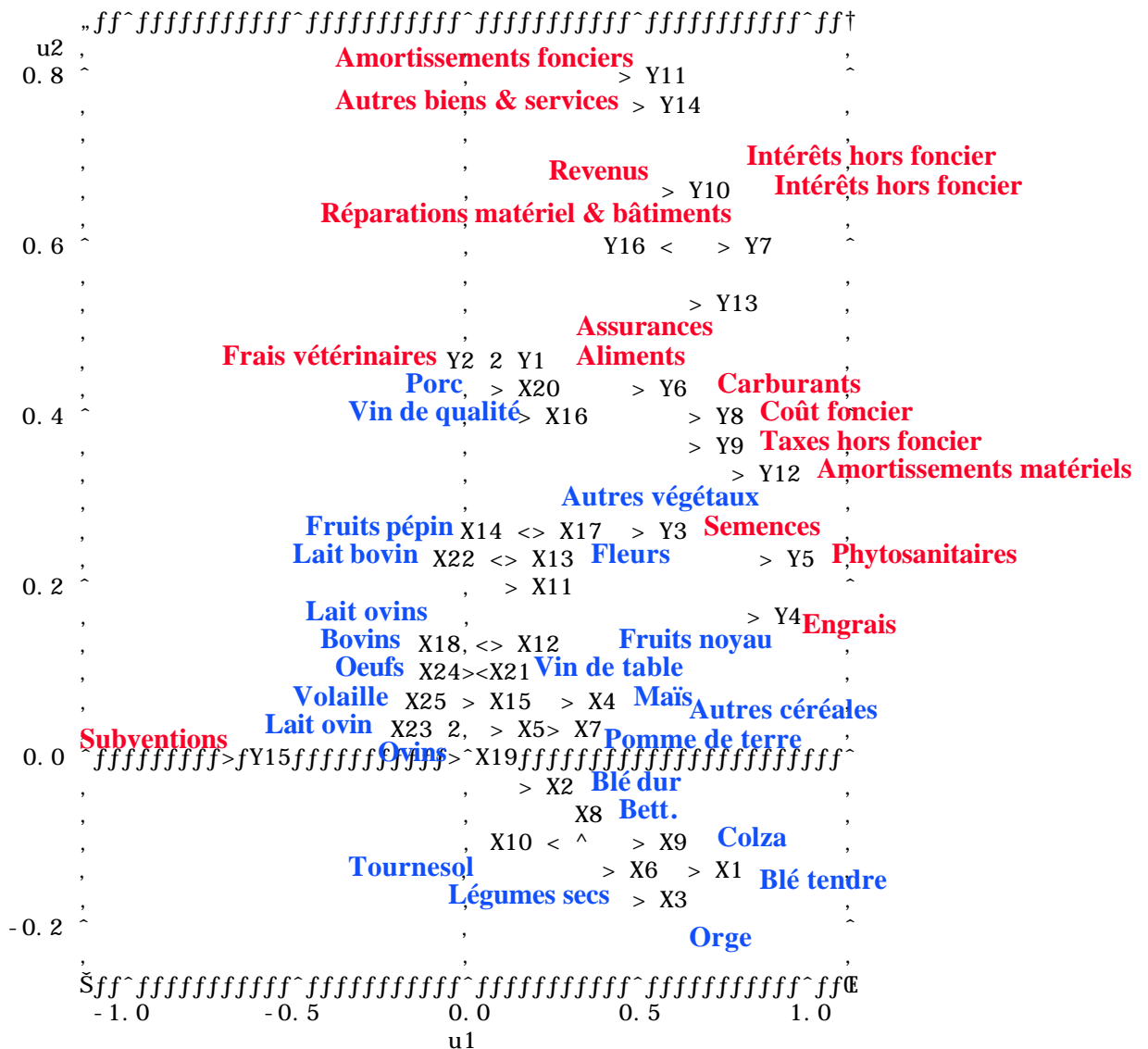
**Figure 9 : Relation entre variables endogènes et variables exogènes dans l'espace des X-pseudocomposantes.**



L'analyse exploratoire peut être complétée par un graphique similaire (figure 10) sur les deux premières Y-pseudocomposantes  $u_1$  et  $u_2$ , où sont également projetées l'ensemble des variables exogènes et endogènes. La première Y-pseudocomposante  $u_1$  oppose clairement les subventions (comptées comme charge négative) aux autres charges. La seconde Y-pseudocomposante  $u_2$  constitue un gradient de charges qui sépare clairement les charges opérationnelles des charges de structure.

**Figure 10 : Relation entre variables endogènes et variables exogènes dans l'espace des Y-pseudo-composantes.**

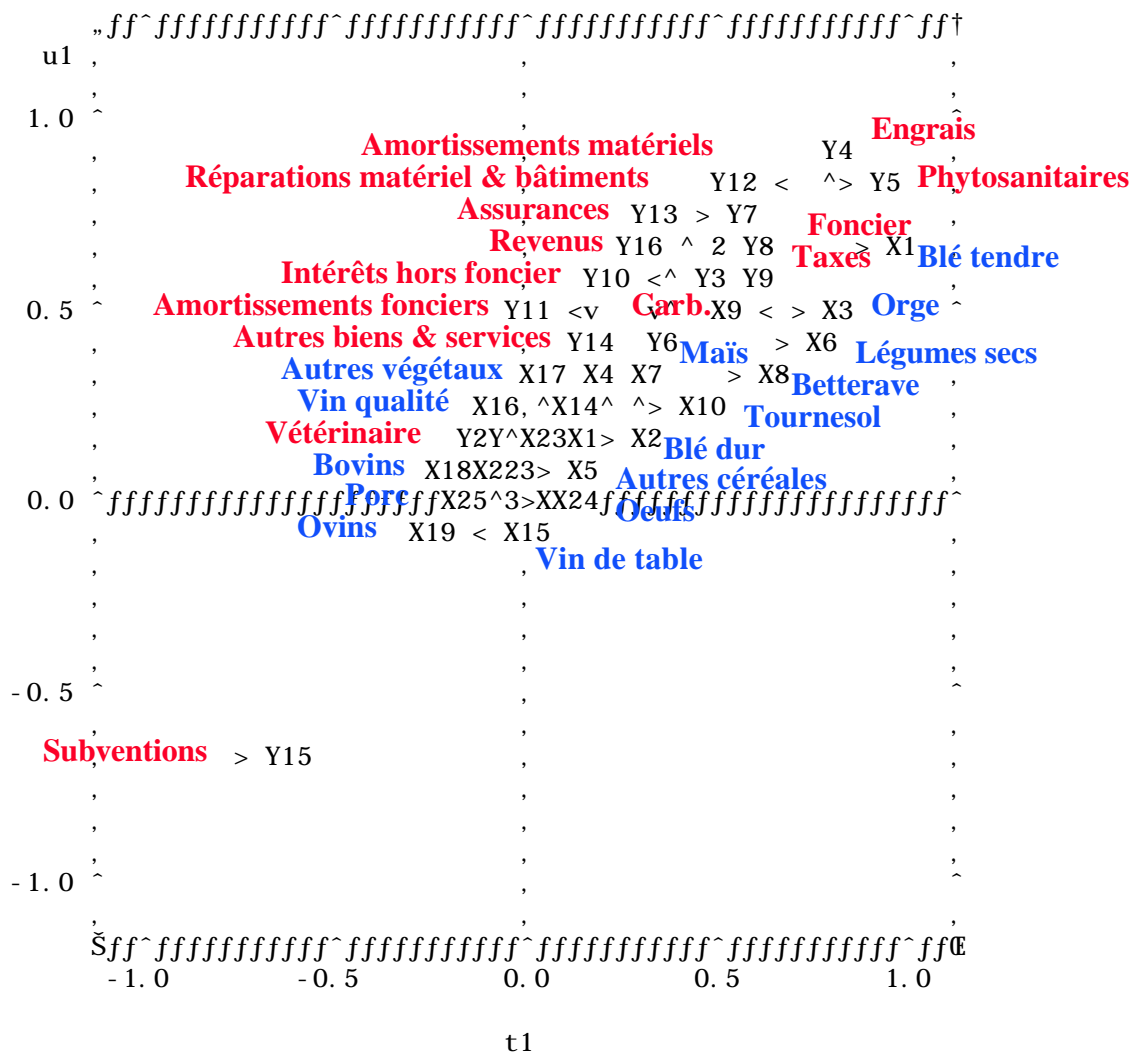
Graphique des variables, corrélations avec  $u_1$  x  $u_2$



L'ajustement principal entre les deux premières pseudo-composantes  $t_1$  et  $u_1$  (figure 11) révèle que l'ensemble des charges de structure se regroupe en haut à gauche du graphique pour s'opposer aux charges spécifiques qui apparaissent liés aux produits de l'autre côté de la diagonale.

**Figure 11 : Ajustement principal des Y-pseudo-composantes et des X-pseudocomposantes..**

Graphique des variables, corrélations avec  $t_1$  x  $u_1$



## 4.4 Comparaison des résultats des deux modèles SURE et PLS

### 4.4.1 Critères d'ajustement statistiques

La comparaison empirique des deux méthodologies d'estimation, estimateurs *SUR* et estimateurs *PLS*, est effectuée sur la base de l'échantillon 1996 du Rica, cette année servant de référence en termes d'indicateur normalisé du potentiel économique des exploitations (rebasement des coefficients de marge brute standard).

Etant donné que les deux modélisations que l'on compare sont issues de cadres théoriques distincts, la difficulté est de trouver des critères de comparaison présentant une définition identique pour les deux méthodes, à défaut une interprétation équivalente :

- Le premier critère proposé, celui de la corrélation  $r_k$  entre les estimations  $\hat{Y}_k$  et les observations des variables endogènes  $Y_k$ , possède une définition identique mais constitue un critère local à chaque équation de charge. On peut pallier cet inconvénient en calculant un indice pour l'ensemble des équations de charge, par exemple la moyenne des coefficients de corrélation sur l'ensemble des  $K$  équations de charge.

**Tableau 4 : corrélation entre charges observées et charges estimées**

Corrélation de $Y_k$ et $\hat{Y}_k$	SUR	PLS
Y1 ALIMENTS	0,960	0,960
Y2 VETERINAIRE	0,871	0,870
Y3 SEMENCES	0,792	0,791
Y4 ENGRAIS	0,879	0,875
Y5 PRODUITS PHYTOSANITAIRES	0,906	0,904
Y6 PRODUITS PETROLIERS	0,745	0,744
Y7 REPARATIONS BAT. & MAT.	0,722	0,721
Y8 COUT FONCIERS	0,796	0,794
Y9 TAXES (HORS FONCIER)	0,642	0,637
Y10 INTERETS (HORS FONCIER)	0,727	0,726
Y11 AMORTISSEMENTS FONC. BAT.	0,735	0,733
Y12 AMORTISSEMENTS MAT.	0,779	0,777
Y13 ASSURANCES	0,736	0,747
Y14 AUTRES BIENS ET SERVICES	0,831	0,829
Y15 AUTRES SUBVENTIONS	0,911	0,911
T16 MARGE NETTE	0,873	0,869
<i>Indice moyen de corrélation</i>	<i>0,807</i>	<i>0,806</i>

On peut vérifier que sur la totalité de l'échantillon du Rica, l'ajustement fourni par les deux types d'estimation est pratiquement équivalent pour chacun des postes de charge, avec un avantage très minime pour l'estimation de type *SUR*. Cette équivalence relève probablement d'un phénomène de convergence asymptotique des estimateurs utilisés, convergence asymptotique établie sur le plan théorique pour les estimateurs *SUR*. On notera que le modèle *PLS* est légèrement meilleur sur le poste « assurances » appartenant à la classe des charges de structure.

- Le second critère choisi est celui du coefficient de détermination (cf. tableau 5). Ce critère est défini pour chaque équation de charge et ne vaut que pour comparer l'estimation des MCO (*OLS-Ordinary Least Squares*) réalisée en préalable de l'estimation *SUR* à celle fournie par la procédure *PLS*. Cependant, on peut comparer sur l'ensemble des équations deux critères globaux ayant une signification équivalente de « pourcentage de variance expliquée » : le  $R^2$  matriciel pondéré (*System weighted R<sup>2</sup>*) pour l'estimation de type *SUR* et le minimum de la racine carrée du *PRESS* moyen (*Minimum root Mean Press*) pour l'estimation de type *PLS*.

**Tableau 5 : comparaison des coefficients de détermination ( $R^2$ )**

<i>Estimations</i> $R^2$ ajusté de $Y_k$	<i>OLS/SUR</i> avec $x_1, \dots, x_{25}$	<i>PLS avec</i> $t_1, \dots, t_{13}$
Y1 ALIMENTS	0,93105	0,9159
Y2 VETERINAIRE	0,81697	0,7441
Y3 SEMENCES	0,70074	0,5634
Y4 ENGRAIS	0,88111	0,7079
Y5 PRODUITS PHYTOSANIRAIRES	0,89169	0,7468
Y6 PRODUITS PETROLIERS	0,67552	0,5257
Y7 REPARATIONS BAT. & MAT.	0,76876	0,4404
Y8 COUT FONCIERS	0,78058	0,6003
Y9 TAXES (HORS FONCIER)	0,63258	0,3708
Y10 INTERETS (HORS FONCIER)	0,68788	0,4911
Y11 AMORTISSEMENTS FONC. BAT.	0,69372	0,4671
Y12 AMORTISSEMENTS MAT.	0,80258	0,5608
Y13 ASSURANCES	0,77778	0,5179
Y14 AUTRES BIENS ET SERVICES	0,79266	0,558
Y15 AUTRES SUBVENTIONS	0,90586	0,8331
T16 REVENU	0,854	0,755
<i>R<sup>2</sup> matriciel</i>	<i>SUR</i>	<i>PLS</i>
<i>System Weighted R-Square</i>	0,7927	
<i>Percent Variation Accounted</i>		0,656

La comparaison globale est légèrement en faveur de l'estimation de type *SUR* mais on observe que pour certains postes de charges fixes, tels que « assurances » et « autres biens et services », le classement des deux modèles s'inverse, la valeur du critère pour l'estimation de type *PLS* étant légèrement meilleure.

- Le troisième critère choisi est celui de l'écart quadratique moyen (**EMQ**) pour chaque équation de charge (cf. tableau 6). Là encore, ce critère est local et ne vaut que pour comparer l'estimation des MCO, (*OLS-Ordinary Least Squares*) réalisée en préalable de l'estimation *SUR*, à celle fournie par la procédure *PLS*. On peut aussi comparer sur l'ensemble des équations deux critères globaux ayant une signification équivalente « d'écart moyen quadratique » : l'écart moyen quadratique matriciel pondéré (*System weighted MSE*) pour l'estimation de type *OLS/SUR* et la racine carrée du PRESS moyen minimum (*Minimum root Mean Press*) pour l'estimation de type *PLS*.

**Tableau 6 : comparaison des écarts moyens quadratiques (EMQ)**

<i>Estimations EMQ de <math>Y_k</math></i>	<i>OLS/SUR</i>	<i>PLS</i>
Y1 ALIMENTS	63513	65679
Y2 VETERINAIRE	14313	14703
Y3 SEMENCES	40829	44151
Y4 ENGRAIS	27994	31503
Y5 PRODUITS PHYTOSANIRAIRES	25868	30607
Y6 PRODUITS PETROLIERS	20440	21034
Y7 REPARATIONS BAT. & MAT.	29000	31011
Y8 COUT FONCIERS	46094	47989
Y9 TAXES (HORS FONCIER)	9789	10093
Y10 INTERETS (HORS FONCIER)	30868	32035
Y11 AMORTISSEMENTS FONC. BAT.	36628	39285
Y12 AMORTISSEMENTS MAT.	52011	54608
Y13 ASSURANCES	15309	14802
Y14 AUTRES BIENS ET SERVICES	100066	119272
Y15 AUTRES SUBVENTIONS	66557	64330
T16 REVENU	156730	158428
<i>EMQ matriciel</i>	<i>SUR</i>	<i>PLS</i>
<i>System Weighted MSE</i>	0,937	
<i>Minimum root Mean Press</i>		0,872

La comparaison globale est également en faveur de l'estimation de type *SUR* même si l'on observe pour le poste des « assurances » et le poste des « subventions », un ajustement légèrement meilleur pour l'estimation de type *PLS*. Remarquons que, là encore, ces deux postes ne sont pas directement liés à la production comme les postes de charges spécifiques comme les engrais ou les aliments.

Notons que ces résultats sont obtenus sur l'ensemble de l'échantillon avec un nombre d'observations qui place l'estimation *SUR* dans une situation relativement favorable où la multicollinéarité n'est pas importante et où l'on les conditions de convergence vers une estimation non biaisée. Les performances de l'estimation *PLS* demeurent bonnes dans cette situation et pratiquement équivalentes à celles de l'estimation *SUR*. Elles s'avèrent parfois meilleurs pour certains postes de charges indirectes. Il reste à comparer les performances des deux modèles sur des estimations régionales où les conditions de convergence asymptotique ne sont pas réunies, d'une part et d'autre part pour des systèmes productifs plus homogènes (combinaison OTEX-CDEX) où la multicollinéarité pourrait se révéler plus importante en raison de l'homogénéisation des contextes productifs et des techniques de production.

#### 4.4.2 Critères de pertinence

Parmi les critères de pertinence des modèles par rapport aux coûts de production que l'on cherche à estimer, on peut sélectionner les suivants :

- les coefficients techniques sont-ils du signe attendu (positif pour les charges, négatif pour les subventions) ?
- le niveau des estimations de ces coefficients techniques est-il conforme aux ordres de grandeur admis par les professionnels ? A défaut de disposer d'une telle expertise qui nécessite de sélectionner un panel représentatif des professionnels pour chaque produit, on comparera les estimations obtenues à celles fournies par les estimateurs précédemment utilisés, MCO et *SUR*, sur un ensemble de produits pour lesquels le modèle Mecpa-ST est couramment utilisé.

On notera que l'estimation *PLS* produit pour l'ensemble des produits des coefficients techniques dont le signe est correct (positif pour les charges, négatif pour les subventions), à l'exception de coefficients dont la valeur absolue est quasi-nulle : charge  $Y_2$  de frais vétérinaires pour le blé tendre, le maïs et la betterave). Ainsi, sans avoir besoin d'imposer une contrainte sur l'ensemble des coefficients pour annuler les coefficients (contrainte imposée dans le modèle *SUR*), le modèle *PLS* affecte un niveau quasi-nul à ce type de dépenses pour la production de céréales.

Cependant, il y a un désaccord important entre le modèle *PLS* et les autres modèles sur le niveau du coefficient technique des subventions ( $Y_{15}$ ) pour la betterave sucrière (cf. figure 14) et le lait (cf. figure 15), productions régulées par des systèmes de quotas de production et dont le soutien ne passe pas ou très peu par des subventions d'exploitation. Dans le cas du lait, cela conduit vraisemblablement à surestimer la marge nette (revenu de l'exploitation). Cette estimation particulière des subventions pourrait refléter en fait l'assolement avec des céréales pour la betterave sucrière et vraisemblablement l'importance prise par les subventions au maïs fourrager pour garantir la marge dans la production laitière. Les autres équations de charges ne présentent pas de désaccord majeur pour l'estimation des coefficients techniques entre les trois méthodes d'estimation.

Ainsi, les particularités de l'estimation de type non paramétrique fournie par la régression *PLS* pourraient se révéler utiles dans l'avenir si l'on souhaite ventiler le paiement unique, décidé en juin 2003, selon les différentes spéculations envisageables sur l'exploitation. Une telle procédure permettrait d'effectuer des simulations sur les choix productifs que pourraient opérer les exploitants agricoles en fonction des signaux du marché. Ces simulations constitueraient un outil pertinent de prévention des situations de crise ou d'instabilité chronique des marchés agricoles pouvant entraîner des préjudices importants pour certains systèmes productifs et désorganiser les filières de l'agro-alimentaire européen.



Figure 12 : Comparaison des estimations MCO, PLS et SUR des coefficients techniques du blé tendre, normalisés pour 1000 unités monétaires de produit brut.

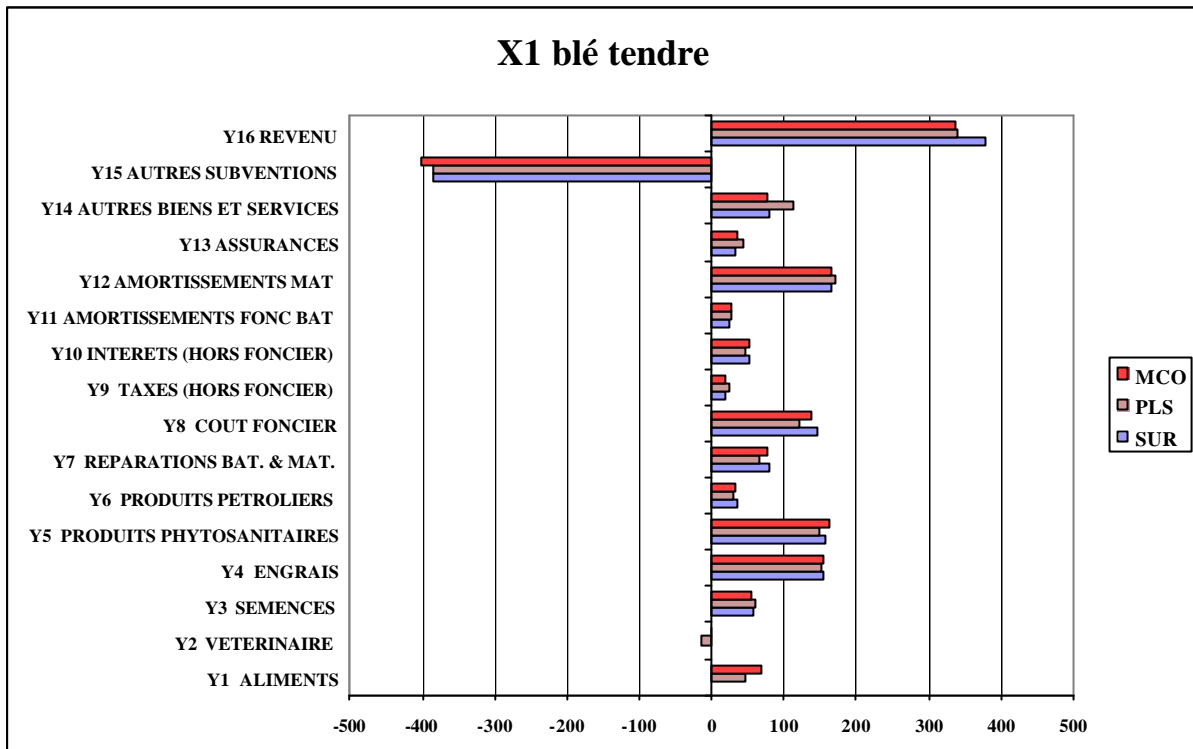
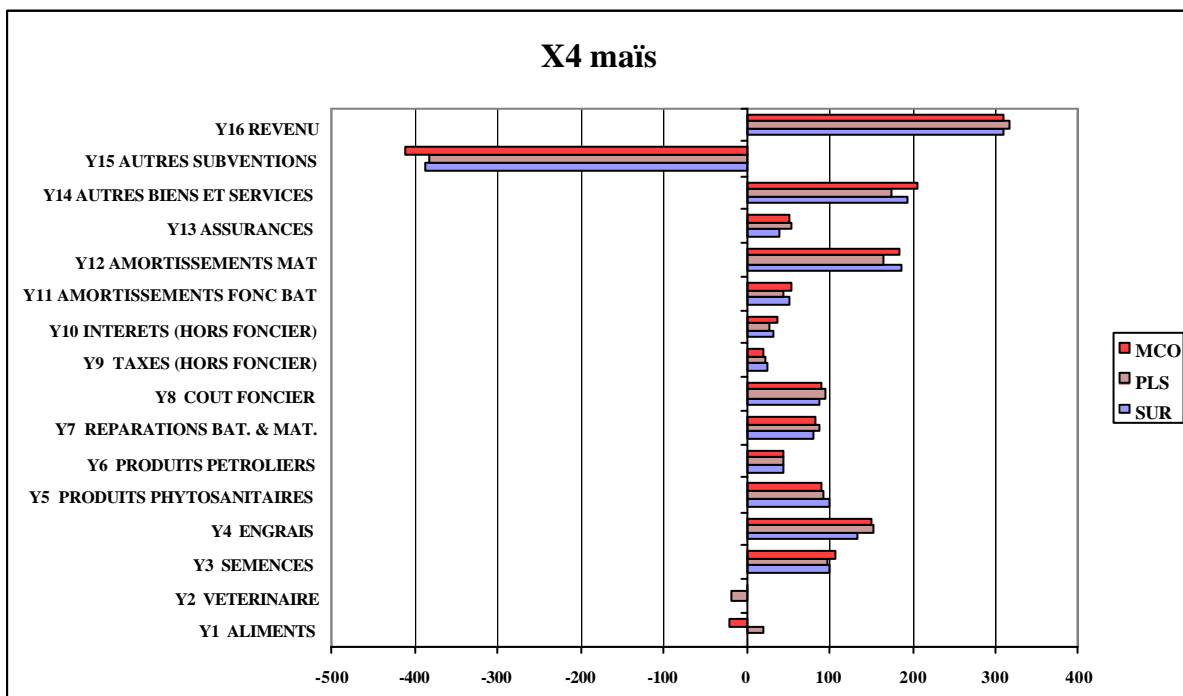
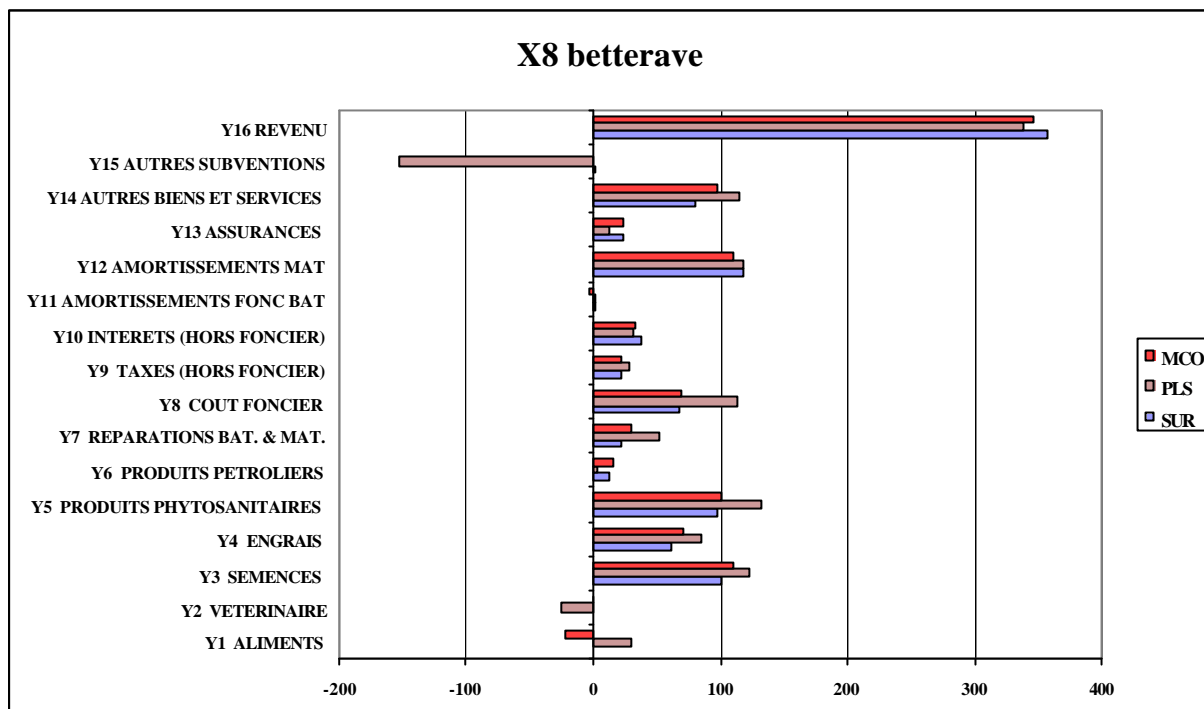


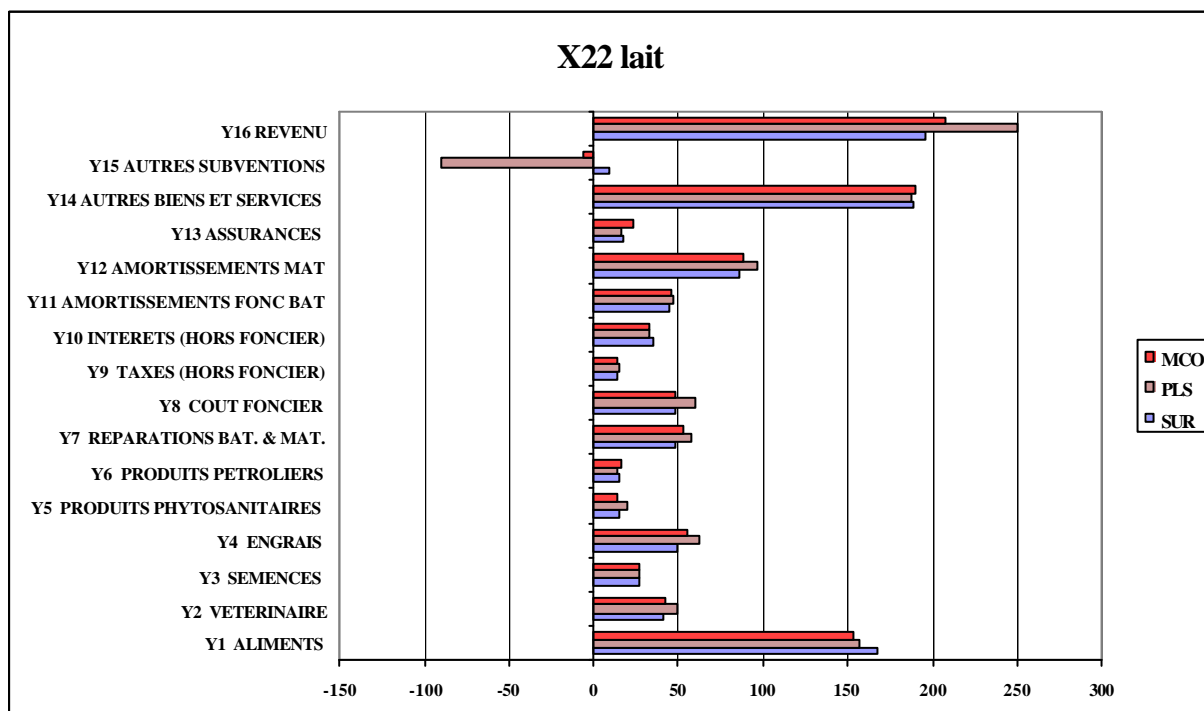
Figure 13 : Comparaison des estimations MCO, PLS et SUR des coefficients techniques du maïs, normalisés pour 1000 unités monétaires de produit brut



**Figure 14 : Comparaison des estimations MCO, PLS et SUR des coefficients techniques de la betterave, normalisés pour 1000 unités monétaires de produit brut.**



**Figure 15 : Comparaison des estimations MCO, PLS et SUR des coefficients techniques du lait, normalisés pour 1000 unités monétaires de produit brut.**



## Annexe 1 : L'algorithme itératif PLS2 de la régression multivariée des moindres carrés partiels (d'après [21]).

### A1.1 Objet de la méthode

La régression PLS (*Partial Least Squares*) ou régression des moindres carrés partiels (MCP) permet de relier un ensemble de variables endogènes  $Y = \{y_1, y_2, \dots, y_p\}$  et un ensemble de variables exogènes  $X = \{x_1, x_2, \dots, x_m\}$ .

L'algorithme de la régression PLS est basé à l'origine sur une méthode itérative, de type *NIPALS* (*Nonlinear Iterative Partial Least Squares*), permettant d'extraire les composantes principales d'un ensemble de variables. L'algorithme est composé de deux étapes : initialisation puis extraction itérative des pseudo-composantes principales (**PCP**).

L'étape d'initialisation consiste à centrer et réduire les variables de chacun des tableaux  $X$  et  $Y$  :

$$E_0 = (I - \mathbf{1}\mathbf{1}'D)XD_{1/s_X} \quad \text{et} \quad F_0 = (I - \mathbf{1}\mathbf{1}'D)YD_{1/s_Y}$$

L'extraction des PCP est basée sur une procédure itérative de recherche de combinaisons linéaires  $t_h$  ( $h^{\text{ième}}$ -PCP en  $X$ ) des colonnes du tableau  $E_0$  (tableau  $X$  centré-réduit), respectivement  $u_h$  ( $h^{\text{ième}}$ -PCP en  $Y$ ) des colonnes du tableau  $F_0$  (tableau  $Y$  centré-réduit) maximisant la covariance du couple de PCP en  $X$  et en  $Y$  à chaque itération  $h$  :

$$\text{cov}(t_h, u_h) = \sqrt{\text{var}(t_h)} \times \sqrt{\text{var}(u_h)} \times \text{cor}(t_h, u_h) = \|t_h\| \times \|u_h\| \times \cos \mathbf{q} \quad \text{où}$$

$\mathbf{q} = \hat{\angle}(t_h, u_h)$  est l'angle entre les deux composantes  $t_h$  et  $u_h$ .

### A1.2 Étapes de l'algorithme

#### A1.2.1 Initialisation

L'étape d'initialisation consiste à centrer et réduire les variables de chacun des tableaux  $X$  et  $Y$

$$E_0 \leftarrow \left( \frac{X - \bar{X}}{s_X} \right)$$
$$F_0 \leftarrow \left( \frac{Y - \bar{Y}}{s_Y} \right)$$

#### A1.2.2 Extraction des $a$ premières composantes MCP

**"Première configuration"**;

Si  $p > m$  alors faire pour  $h = 1, \dots, a$  ;

"Si  $p$  le nombre de variables expliquées est supérieur à  $m$  nombre de variables explicatives, on diagonalise une matrice  $(m, m)$ "

**{début**

$$l = h - 1 ;$$

"Calcul de la matrice de covariance  $V_k$  " ;

$$V_l \leftarrow \frac{1}{n-1} (F_l' E_l) ;$$

"Analyse en composantes principales de  $V_l V_l'$  " ;

$$V_l V_l' = \sum_{j=1}^m \mathbf{I}_j v_j v_j'$$

Extraction de  $\boxed{v_h}$  vecteur propre normé associé à la plus grande valeur propre  $\boxed{\mathbf{I}_h}$  de

$$V_l V_l' = \frac{1}{p} E_l' F_l F_l' E_l ;$$

"Calcul de la  $h^e$  X -composante principale de  $E_0$  " ;

$$t_h \leftarrow (E_l v_h) ;$$

Calcul de  $w_h$  vecteur propre associé à la plus grande valeur propre  $\mathbf{I}_h$  de  $V_l V_l' = F_l' E_l E_l' F_l$  ;

$$\boxed{w_h \leftarrow \frac{1}{\sqrt{\mathbf{I}_h p}} (V_k v_h) ;}$$

"Calcul de la  $h^e$  Y -composante MCP de  $F_0$  " ;

$$\boxed{u_h \leftarrow (F_l w_h) ;}$$

"Régression MCO de  $E_l$  sur  $t_h$  :  $E_l = t_h p_h' + E_h$  " ;

$$p_h ; \quad \text{"Calcul de la } h^e \text{ X -saturation"} ;$$

$$\boxed{E_h \leftarrow E_l - t_h p_h' ;} \quad \text{"Calcul du } h^e \text{ X -résidu"} ;$$

"Régression MCO de  $F_l$  sur  $t_h$  :  $F_l = t_h r_h' + F_h$  " ;

$$\boxed{F_h \leftarrow F_l - t_h r_h' ;} \quad \text{"Calcul du } h^e \text{ Y -résidu"} ;$$

**fin** ;

**"Seconde configuration"** ;

Sinon alors faire pour  $h = 1, \dots, a$  ;

"Calcul pour  $m \geq p$ , si  $m$  le nombre de variables explicatives est supérieur à  $p$ , le nombre de variables expliquées, on diagonalise une matrice  $(p, p)$  " ;

**{début**

$$l = h - 1 ;$$

"Calcul de la matrice de covariance  $V_l$  " ;

$$V_l \leftarrow \frac{1}{n-1} (E_l' F_l) ;$$

"Analyse en composantes principales de  $V_k V_k'$  " ;

$$V_l V_l' = \sum_{k=1}^p \mathbf{I}_k w_k w_k'$$

Extraction de  $\boxed{w_h}$  vecteur propre normé associé à la plus grande valeur propre  $\boxed{\mathbf{I}_h}$  de

$$V_l V_l' = \frac{1}{m} F_l' E_l E_l' F_l ;$$

"Calcul de la  $h^e$  Y-composante MCP de  $F_0$  " ;

$$u_h \leftarrow (F_l w_h) ;$$

Calcul de  $v_l$  vecteur propre associé à la plus grande valeur propre  $I_{1_l}$  de  $V_h V_h' = E_h' F_h F_h' E_h$  ;

$$v_h \leftarrow \frac{1}{\sqrt{I_{1_l} m}} (V_l w_h) ;$$

"Calcul de la  $h^e$  X-composante MCP de  $E_0$  " ;

$$t_h \leftarrow (E_l v_h) ;$$

"Régression MCO de  $E_l$  sur  $t_h$  :  $E_l = t_h p_h' + E_h$  " ;

$$p_h ; \quad \text{"Calcul de la } h^e \text{ X-saturation"} ;$$
$$E_h \leftarrow E_h - t_h p_h' ; \quad \text{"Calcul du } h^e \text{ X-résidu"} ;$$

"Régression MCO de  $F_l$  sur  $t_h$  :  $F_l = t_h r_h' + F_h$  " ;

$$F_h \leftarrow F_l - t_h r_h' ;$$

**fin** ;

## Bibliographie

- [ 1] Acher O. « Blé tendre : coûts de production 2001 », Office national interprofessionnel des céréales, 4 p., juillet 2002.
- [2] Aufrant M. « Les coûts de production des grands produits agricoles : évolution de 1970 à 1978 », Institut national de la statistique et des études économiques, *Archives et Documents* n°64, 246 p., janvier 1983.
- [3] Aufrant M., Picard H. « Les parités de pouvoir d'achat entre pays », *Le Courrier des Statistiques* n°44, pp. 8-16, octobre 1987.
- [4] Butault J-P., Carles R., Hassan D, Reignier E. « Les coûts de production des principaux produits agricoles dans la Communauté européenne », Office des publications officielles des Communautés européennes, 293 p., 1988.
- [5] Butault J-P., Carles R., Hassan D, Reignier E. « L'agriculture dans la CEE. Tome 3. Coûts de production », *Les collections de l'INSEE, série Entreprise*, n°112, 141 p., septembre 1988.
- [6] Butault J-P., Carles R., Hassan D, Reignier E. « Le coût de production des divers produits agricoles dans la CEE », *Les collections de l'INSEE, série Archives* n°260, 124 p., octobre 1988.
- [7] Butault J-P., Cyncynatus M. (dir.) « Les coûts de production des principaux produits agricoles de la Communauté européenne en 1984-85-86 (RICA). Fascicule 1 : Présentation du modèle et des principaux résultats nationaux. Fascicule 2 : Études transversales. Fascicule 3 : Analyses par produit : partie A produits végétaux ; partie B produits animaux », Commission européenne, Bruxelles, 367 p., septembre 1990.
- [8] Butault J-P., Cyncynatus M. (dir.) « Coûts de production et compétitivité des agricultures européennes ». Institut national de la recherche agronomique, Département Économie et Sociologie Rurales, série *Actes et Communications* n°5, 248 p., 1991.
- [ 9] Butault J-P., Delame N., Rouselle J-M. « La productivité dans l'agriculture française varie beaucoup selon les produits », Institut national de la recherche agronomique, *Inra Sciences Sociales* n°3, 4p., juin 1995.
- [10] Chantry E. « Le Réseau d'information comptable agricole : un trentenaire en pleine jeunesse », Institut national de la statistique et des études économiques, *Courrier des statistiques* n°85-86, pp. 21-26, 1998.
- [11] Chantry E. « Le Réseau d'information comptable agricole (RICA) : un outil unique de connaissance des agricultures européennes », Ministère de l'Agriculture, de l'Alimentation, de la Pêche et des Affaires rurales, *Notes et études économiques* n°18, pp. 9-19, avril 2003.
- [12] Corlieu (de) T., Desbois D., Pollet P. « Blé tendre, pomme de terre, porc : baisse des coûts de production entre 1990 et 1997 », ministère de l'Agriculture et de la Pêche, Agreste, *Les Cahiers* n°40, pp. 29-34, juin 1999.
- [13] Desbois D., Legris B. « D'une réforme de la politique agricole commune à l'autre. Baisse des prix et maîtrise des coûts », Institut national de la statistique et des études économiques, *Insee Première* n°927, 4 p., 16 octobre 2003.
- [14] Desbois D., Pollet P. « Evaluation microéconomique de la marge brute standard sur la base du RICA », Institut national de la statistique et des études économiques, *Insee Méthodes* n°101 (tome 2), Journées de méthodologie statistique - décembre 2000, pp. 395-429, 2002.
- [15] Divay JF., Meunier F. « Deux méthodes de confection du tableau entrées-sorties ». Institut national de la statistique et des études économiques, *Annales de l'Insee* n°37, pp. 59-108, janvier-mars 1980.
- [16] Jong (de), S. « SIMPLS: An Alternative Approach to Partial Least Squares Regression », *Chemometrics and Intelligent Laboratory Systems*, 18, pp. 251-263, 1993.
- [17] Pollet P. « Quinze années de grandes cultures : baisse des prix et réduction des coûts », Institut national de la statistique et des études économiques, *Insee-Première* n°473, 4 p., juillet 1996.

- [18] Pollet P. « Le modèle sur les coûts de production agricole », avec la collaboration de J-P. Butault (Inra) et E. Chantry (SCEES), Division agriculture de l'INSEE, *Document de travail* n°E9802, 72 p., 1998.
- [19] Pollet P. « Du rendement à la rentabilité en grandes cultures », Institut national de la statistique et des études économiques, *Économie et statistique* n°329-330, pp. 127-146, septembre-octobre 1999.
- [20] SCEES « Les coefficients de marge brute standard 1996 », ministère de l'Agriculture et de la Pêche, *Les Cahiers de l'Agreste hors série* n°1, 51p., janvier 2001.
- [21] Tenenhaus, M., Gauchi J.-P., Ménardo C. « Régression PLS et applications », *Revue de Statistique Appliquée*, XLIII (1), pp. 7-63, 1995.
- [22] Tenenhaus, M. *La régression PLS. Théorie et pratique*, Éditions Technip, 254 p., 1998.
- [23] Tobias, R.D. « An Introduction to Partial Least Squares Regression », SAS Institute, Cary, NC, 8 p., 1996.
- [24] Voet (van der), H. « Comparing the Predictive Accuracy of Models Using a Simple Randomization Test », *Chemometrics and Intelligent Laboratory Systems*, 25, pp. 313-323, 1994.
- [25] Wold, H. « Estimation of Principal Components and Related Models by Iterative Least Squares », in *Multivariate Analysis*, ed. P. R. Krishnaiah, New York: Academic Press, pp. 391-420, 1966.
- [26] Zellner, A. « An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias », *Journal of the American Statistical Association* n° 57, pp. 348-368, 1962.