



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Frontière de production
et données de panel :
applications aux
exploitations céréalières

Isabelle PERRIGNE

Production frontier and panel data : application to grain farmers

Key-words :
panel data econometrics,
production frontier,
technical efficiency

Frontière de production et données de panel : applications aux exploitations céréalières

Mots-clés :
économétrie des données de panel, frontière de production, efficacité technique

Summary – *This paper deals with the issue of estimating a stochastic production frontier and of measuring technical efficiency from a complete panel of grain farmer from 1982 to 1986.*

A survey of classical panel data econometric methods (Within and GLS) and their applications underline the limits of their use with respect to the empirical results and the characteristics of our data set. Moreover, the Hausman test shows a correlation between inputs and the error term which represents inefficiency, which invalidates GLS estimation.

The use of a new method proposed by Ivaldi, Perrigne and Simioni (1994) based on a covariance structural analysis allows to reveal correlations between regressors and the individual effect as well as to estimate an individual time-varying technical efficiency. Empirical evidences show a negative correlation between materials and efficiency and a declining trend for efficiency between 1982 and 1986.

Résumé – Ce papier propose d'estimer une frontière de production stochastique et de mesurer l'efficacité technique à partir d'un échantillon de panel cylindré d'exploitations céréalières de 1982 à 1986 en utilisant diverses méthodes.

Une présentation des méthodes usuelles de données de panel (*Within* et *GLS*) et leurs applications mettent en lumière les limites de celles-ci quant aux résultats obtenus et aux spécificités de l'échantillon. De plus, le test d'Hausman montre l'existence d'une possible corrélation entre les inputs utilisés et le terme d'erreur représentant l'inefficacité, ce qui rend caduque l'estimation *GLS*.

L'utilisation d'une méthode récente proposée par Ivaldi, Perrigne et Simioni (1994), basée sur une modélisation en termes d'analyse structurelle des covariances permet de révéler les corrélations entre les régresseurs et l'effet individuel. Cette modélisation permet aussi d'estimer l'efficacité technique de chaque individu comme une fonction du temps.

Les résultats empiriques montrent d'une part, l'existence d'une corrélation négative entre les consommations intermédiaires et l'efficacité et d'autre part, une décroissance des niveaux d'efficacité entre 1982 et 1986.

* Station d'économie et sociologie rurales de l'INRA, Auzeville, BP 27, 31326 Castanet Tolosan Cedex.

LES modèles de frontière de production ont été l'objet de nombreuses études empiriques en économie agricole (voir Battese, 1992 pour un tour d'horizon de cette littérature). Il est en effet intéressant d'étudier comment se situent les agriculteurs en terme d'efficacité technique dans un secteur où le travail a une forte productivité et où l'on constate une importante substitution capital-travail, et de déterminer ensuite les facteurs expliquant les différences observées entre ces agriculteurs.

Cet article* présente diverses méthodes d'estimation de frontière de production et de mesures d'efficacité technique à partir d'un échantillon de panel. Nous les illustrerons ensuite sur des données agricoles du RICA.

La modélisation de frontière stochastique de production a été initialement proposée par Aigner, Lovell et Sickles (1977) et repose sur la décomposition du terme d'erreur en deux parties. Tout d'abord l'introduction d'un terme aléatoire non borné dans la fonction frontière permet la prise en compte d'erreurs de mesure, de spécification et d'aléas pouvant affecter le processus de production. Ces derniers peuvent être non négligeables en agriculture: c'est le cas notamment des aléas climatiques. L'autre terme représente l'inefficacité. Cette modélisation a néanmoins deux inconvénients majeurs. D'une part, elle nécessite des hypothèses a priori sur la distribution des deux termes d'erreur. D'autre part, elle ne permet pas d'identifier pour chaque individu la décomposition du terme d'erreur. Il est donc impossible de calculer des mesures individuelles d'efficacité technique.

La généralisation aux données de panel avec utilisation de méthodes classiques de données de panel telles que *Within* et *GLS* présente plusieurs avantages: aucune hypothèse sur la distribution du terme d'erreur n'est exigée et on peut obtenir des mesures individuelles d'efficacité (voir Schmidt et Sickles, 1984). Pour ces raisons, de nombreux auteurs, dont Schmidt (1986), ont encouragé l'utilisation de données de panel pour estimer des frontières de production et des mesures d'efficacité technique. Plusieurs recherches ont été effectuées dans cette voie. Ces méthodes simples à mettre en œuvre présentent cependant des inconvénients pour être utilisées efficacement. D'une part, elles demandent un grand nombre d'individus et/ou de périodes dans l'échantillon de panel. Or les données agricoles ne permettent pas en général de construire des échantillons de panel avec un nombre substantiel de périodes du fait de la périodicité annuelle des données. Ainsi, les études empiriques en économie agricole dé-

* Cet article fait suite aux travaux de recherche effectués pour la thèse de doctorat. L'auteur remercie Marc Ivaldi et Michel Simioni pour les conseils qui lui ont été prodigués tout au long de l'encadrement de sa thèse, ainsi que Quang Vuong et deux lecteurs anonymes pour leurs commentaires.

nombrées par Battese (1992) portent sur des échantillons variant de 3 à 10 périodes. De plus, le modèle à effets aléatoires (méthode *GLS*) suppose l'indépendance entre les régresseurs et le terme d'erreur. Or dans le contexte d'estimation de frontière de production et de mesure d'efficacité technique, un individu connaît a priori son niveau d'efficacité technique et peut ainsi modifier en conséquence ses décisions et choix technologiques. Il paraît alors raisonnable de supposer qu'il existe une éventuelle corrélation entre les quantités d'inputs et le niveau d'efficacité. Le test d'Hausman (1978) permet de tester l'existence de cette corrélation en comparant les résultats entre les deux méthodes *Within* et *GLS* (la première n'exigeant aucune hypothèse a priori, la seconde supposant une hypothèse d'indépendance entre les régresseurs et l'effet individuel).

Une méthode économétrique classique tenant compte de l'existence de cette corrélation est l'emploi de variables instrumentales. Celle-ci suppose de disposer d'un nombre conséquent d'instruments qui sont des variables exogènes exclues préalablement du modèle. Une seconde méthode a été proposée par Hausman et Taylor (1981) dans le cadre d'estimation d'effets individuels non observables où les variables non corrélées avec l'effet individuel sont considérées comme des instruments. Il est à noter que les auteurs proposent un test qui permet de tester *ex post* l'a priori sur les corrélations. Cornwell, Schmidt et Sickles (1990), utilisant la méthode des variables instrumentales, ont proposé l'estimation de mesures d'efficacité comme fonctions quadratiques du temps. Néanmoins, l'utilisation de cette méthode exige d'avoir un a priori sur le régresseur effectivement corrélé avec l'efficacité. De plus, l'estimation convergente de mesures individuelles d'efficacité demande dans le modèle de Cornwell *et al.* un nombre substantiel de périodes qui ne sont pas toujours disponibles dans des données de panel.

Suite à ces difficultés, Ivaldi, Perrigne et Simioni (1994) ont proposé de réécrire un modèle de frontière stochastique de production dans un cadre d'analyse structurelle des covariances définie par Jöreskog (1970, 1977). En traitant les effets individuels inconnus comme des variables latentes, cette méthode d'estimation permet aux données de révéler d'éventuelles corrélations entre les quantités d'inputs et les effets individuels. Les effets individuels peuvent être estimés et donc les mesures d'efficacité retrouvées par ce que l'on appelle des facteurs scores, c'est-à-dire les valeurs calculées des variables latentes à partir des variables observées et des paramètres estimés. De plus, ces mesures d'efficacité peuvent varier dans le temps.

Nous présenterons ici brièvement les caractéristiques de chacune des méthodologies discutées précédemment (*Within*, *GLS*, *Analyse structurelle des covariances*). Nous ferons abstraction de la méthode classique des variables instrumentales car nous ne disposons pas d'instruments pertinents. Quant à l'application de la méthode proposée par Hausman et Taylor (1981), bien qu'un test soit possible, nous n'avons pas d'a priori

assez fort sur le ou les inputs corrélés avec le terme d'efficacité lors des applications. Il est à noter qu'avec une fonction de production à trois inputs, six tests seraient nécessaires. La seconde partie de l'article est consacrée à l'application de ces méthodes pour estimer une frontière de production et les niveaux individuels d'efficacité à partir d'un échantillon d'exploitations céréalières issu du RICA entre 1982 et 1986.

MÉTHODOLOGIE

Cette section présente les principales caractéristiques et étapes des méthodes d'estimation mentionnées dans l'introduction. Pour une présentation plus détaillée des méthodes *Within* et *GLS*, le lecteur peut consulter Hsiao (1986) et Schmidt et Sickles (1984). Notons que cette section est exclusivement consacrée au panel cylindré qui se définit comme un échantillon fixe d'individus dont on observe certaines variables durant un certain nombre de périodes.

Le modèle à effets fixes

Il existe deux sources de variabilité provenant des dimensions individuelle et temporelle des données. Ces deux notions de variabilité sont à l'origine de deux méthodes d'estimation: la méthode *Within* (ou "intra") qui privilégie la dimension temporelle et la méthode *Between* (ou "inter") qui privilégie la dimension individuelle. Nous présentons ici la méthode *Within*.

Considérons un échantillon de panel composé de N individus sur T périodes. Nous proposons d'estimer le modèle à effets fixes suivant:

$$\begin{cases} y_{it} = X_{it}\beta + u_{it}, \\ u_{it} = -\alpha_i + \varepsilon_{it}, \end{cases} \quad i = 1, \dots, N \text{ et } t = 1, \dots, T, \quad (1)$$

où α_i est l'effet individuel s'interprétant comme le terme mesurant l'efficacité d'un individu, et où ε_{it} est un bruit statistique. X_{it} et β sont de dimension respective $(1 \times K)$ et $(K \times 1)$. Nous empilons les T équations (1) pour chaque individu puis les N individus, le modèle s'écrit sous la forme matricielle:

$$y = X\beta + u \quad (2)$$

où y , X , et u sont de dimension respective $(NT \times 1)$, $(NT \times K)$, et $(NT \times 1)$. Nous faisons les hypothèses usuelles $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sigma_\varepsilon^2(I_N \otimes I_T)$. L'opérateur *Within* se définit de la façon suivante. Posons:

$$W = I_N \otimes (I_T - J_T/T) \quad (3)$$

où I_T est la matrice identité de dimension $(T \times T)$ et J_T une matrice constituée uniquement de 1 de dimension $(T \times T)$. Cet opérateur réalise

la transformation $I_T - J_T/T$ sur les données, c'est-à-dire le calcul des écarts par rapport aux moyennes individuelles. En prémultipliant par W , le modèle (2) s'écrit $Wy = WX\beta + Wu$.

En estimant cette dernière équation par les moindres carrés linéaires nous obtenons l'estimateur *Within*:

$$\hat{\beta}_W = (X'WX)^{-1} X'Wy \quad (\text{car } W \text{ est idempotente}) \quad (4)$$

Cet estimateur est convergent lorsque le produit $N T$ tend vers l'infini.

La méthode *Within* permet d'estimer les effets individuels et par conséquent elle peut fournir une mesure individuelle d'efficacité dans le cadre de l'estimation d'une frontière de production. Il est à noter cependant que l'estimation de l'effet individuel n'est convergente que lorsque T tend vers l'infini.

Soit une fonction de production log-linéaire de type Cobb-Douglas:

$$y_{it} = \beta_0 + X'_{it}\beta + v_{it}, \quad i = 1, \dots, N \text{ et } t = 1, \dots, T \quad (5)$$

où $v_{it} = -\alpha_i + \varepsilon_{it}$. L'estimateur *Within* ne permet pas d'estimer la constante β_0 , nous réécrivons l'équation précédente sous la forme:

$$y_{it} = X'_{it}\beta + \gamma_i + \varepsilon_{it}, \quad \text{où } \gamma_i = \beta_0 - \alpha_i \quad (6)$$

En appliquant l'opérateur *Within* sur cette équation, les moindres carrés ordinaires donnent une estimation de β . Les γ_i s'obtiennent en posant $\hat{\gamma}_i = y_{i.} - \hat{\beta}_W X_{i.}$ où $y_{i.}$ est la moyenne des T y_{it} observations pour le i ème individu et $X_{i.}$ est un vecteur de dimension $(K \times 1)$ des moyennes des $T x_{itk}$ observations pour le i ème individu. Nous effectuons une translation de la fonction afin d'obtenir des valeurs α_i positives et d'identifier β_0 .

Ainsi, $\hat{\alpha}_i = \max_j \hat{\gamma}_j - \hat{\gamma}_i$ et $\hat{\beta}_0 = \max_j \hat{\gamma}_j$ d'où la mesure d'efficacité technique dans le cadre d'une fonction de production log-linéaire de type Cobb-Douglas:

$$TE_i = \exp(-\hat{\alpha}_i) \times 100$$

Dans le cadre d'estimation d'une frontière de production et de mesures individuelles d'efficacité technique, cette méthode présente un inconvénient majeur: elle ne permet pas une estimation convergente des effets individuels si le nombre de périodes de l'échantillon est petit. Or, la périodicité des données sur les exploitations agricoles est en général annuelle, ce qui limite la dimension temporelle de ces données.

Le modèle à effets aléatoires

Le modèle que nous cherchons à estimer s'écrit sous la forme:

$$y_{it} = X_{it}\beta + u_{it}, \quad \text{avec } u_{it} = -\alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (7)$$

sous les hypothèses suivantes :

$$\begin{aligned} E(u_{it} u_{it}') &= \begin{cases} \sigma_\alpha^2 & , \text{ si } t \neq t' \\ \sigma_\alpha^2 + \sigma_\varepsilon^2 & , \text{ si } t = t' \end{cases} \\ E(u_{it} u_{i't'}) &= 0 & \text{ si } i \neq i' \\ E(\alpha_i | x_i) &= 0 \end{aligned} \quad (8)$$

d'où $E(uu') = \Omega \neq \sigma_u^2 I_{NT}$ avec $\sigma_u^2 = \text{Var}(u_{it})$.

Nous nous situons dans le cadre d'application des moindres carrés généralisés qui donnent le meilleur estimateur sans biais de variance minimale suivant :

$$\hat{\beta}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \quad (9)$$

Cet estimateur est convergent lorsque le produit $N T$ tend vers l'infini. Il est à noter que cet estimateur suppose l'indépendance entre l'effet individuel et les régresseurs.

Nous ne connaissons pas σ_α^2 et σ_ε^2 . Une étude de la forme de la matrice de covariance Ω va permettre de déduire des estimations pour ces deux variances inconnues.

La matrice Ω peut s'écrire sous la forme $\Omega = I_N \otimes \Sigma$, où $\Sigma = \sigma_\alpha^2 J_T + \sigma_\varepsilon^2 I_T$. Suite à une décomposition de la matrice Σ , nous obtenons :

$$\Omega = I_N \otimes \Sigma = \sigma_\varepsilon^2 \left(W + \frac{1}{\theta^2} B \right) \quad (10)$$

où $B = (T_N \otimes J_T / T)$ est l'opérateur *Between*⁽¹⁾ et $\theta^2 = \sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + T \sigma_\alpha^2)$.

Le paramètre θ^2 peut s'interpréter à un facteur $1/(T-1)$ près comme le rapport de la variance des résidus *Within* sur la variance des résidus *Between*. Ainsi, nous pouvons estimer θ^2 à partir des résidus des régressions *Within* et *Between* et en déduire des estimations σ_α^2 et σ_ε^2 ⁽²⁾.

Nous définissons alors l'estimateur des moindres carrés quasi généralisés (*QGLS*) :

$$\hat{\beta}_{QGLS} = (X' \hat{\Omega}^{-1} X)^{-1} X' \hat{\Omega}^{-1} y \quad (11)$$

avec $\hat{\Omega} = \hat{\sigma}_\varepsilon^2 (W + B/\hat{\theta}^2)$.

Dans le cas de l'estimation d'une frontière de production à partir d'une fonction de type Cobb-Douglas, nous effectuons les mêmes étapes de calcul que dans le modèle à effets fixes pour obtenir l'estimation des effets individuels α_i et de la constante β_0 c'est-à-dire une translation de

⁽¹⁾ L'opérateur B est tel que $B \times W = 0$ et $B + W = 1$. La méthode *Between* s'applique de la même façon que la méthode *Within*.

⁽²⁾ La variance des résidus de la régression *Within* est égale à $\sigma_\varepsilon^2(1 - 1/T)$ et celle des résidus de la régression *Between* est égale à $\sigma_\alpha^2 + \sigma_\varepsilon^2/T$.

la fonction de production (pour un exposé détaillé voir Perrigne, 1991). Les mesures individuelles d'efficacité s'écrivent sous la même forme.

Il est intéressant d'étudier le comportement asymptotique de l'estimateur *GLS* et ainsi de comparer les performances asymptotiques des estimateurs *Within* et *GLS*. Quand T tend vers l'infini, les deux estimateurs sont asymptotiquement équivalents car θ^2 tend vers 0. Lorsque T est fixé, il n'y a plus d'équivalence asymptotique entre les deux estimateurs. Nous pouvons démontrer que la variance de l'estimateur obtenu par *QGLS* est inférieure à celle de l'estimateur obtenu par *Within*. Il semble donc qu'en présence d'un nombre limité de périodes T l'estimateur *QGLS* soit plus performant. Ce résultat n'est valable que si les hypothèses sous-jacentes du modèle à effets aléatoires sont respectées en particulier la non-corrélation entre les régresseurs et le terme d'erreur. Cette dernière hypothèse ne semble pas réaliste dans l'estimation de frontière de production. En effet, l'output de chaque individu peut être affecté par sa capacité de "management" inobservable qui est elle-même liée aux quantités d'inputs utilisées selon la diffusion du progrès technique dans le secteur considéré. Ce dernier phénomène est fréquemment observé en économie de la production⁽³⁾.

Analyse structurelle des covariances

La discussion précédente nous incite à chercher d'autres modèles pour à la fois estimer de façon convergente les effets individuels lorsque T est petit et tenir compte d'une éventuelle corrélation entre ces derniers et les régresseurs. Nous présentons dans cette section la méthode proposée par Ivaldi, Perrigne et Simioni (1994) pour estimer une frontière de production et des mesures individuelles d'efficacité variant dans le temps. Cette méthode repose sur l'analyse structurelle des covariances et prend en compte l'existence éventuelle de corrélations entre les régresseurs et l'effet individuel.

Le modèle

Nous pouvons écrire un modèle de frontière de production dans le cadre d'analyse structurelle des covariances. L'objectif est de tenir compte d'une possible corrélation entre le terme d'erreur et les variables exogènes. Nous considérons alors l'effet individuel comme une variable

⁽³⁾ Cette discussion pose le problème du choix entre deux modélisations selon que les effets sont considérés comme fixes ou aléatoires. Ainsi que le fait remarquer Hsiao (1986), les estimations des deux modèles donnent des résultats très différents lorsque T est relativement petit et N grand. Il serait important, selon l'auteur, de privilégier l'utilisation des variations des données entre individus. Il n'existe néanmoins pas de règle générale de choix entre les deux modèles.

latente. Ce dernier peut s'écrire alors comme une fonction de variables exogènes. Reprenant la formulation de Cornwell, Schmidt et Sickles (1990), ces variables exogènes w_t sont des variables temporelles indépendantes des individus. Le modèle s'écrit :

$$\begin{cases} y_{it} = \beta x_{it} + v_{it} \\ v_{it} = w_t \theta_i + \varepsilon_{it} \end{cases} \quad \forall i = 1, \dots, N \text{ et } t = 1, \dots, T, \quad (12)$$

où x_{it} est un vecteur de variables exogènes de dimension $(K \times 1)$, y_{it} est la variable endogène, β' est un vecteur de paramètres de dimension $(K \times 1)$, w_t' est un vecteur de variables exogènes variant dans le temps de dimension $(L \times 1)$, θ_i est un vecteur de paramètres individuels de dimension $(L \times 1)$ appelés les effets individuels et ε_{it} est un terme d'erreur d'espérance nulle et de variance σ^2 . Soit Δ la matrice de covariance de θ_i .

Dans le cadre d'analyse structurelle des covariances et en considérant un modèle multivarié empilant les T équations pour chaque individu, le modèle s'écrit :

$$\begin{pmatrix} y_i \\ v_i \end{pmatrix} = \begin{pmatrix} 0_{TxT} & I_{TxT} \\ 0_{TxT} & 0_{TxKT} \end{pmatrix} \begin{pmatrix} y_i \\ v_i \end{pmatrix} + \begin{pmatrix} I_{TxT} \otimes \beta & 0_{TxL} \\ 0_{TxKT} & W \end{pmatrix} \begin{pmatrix} x_i \\ \theta_i \end{pmatrix} + \begin{pmatrix} 0_{Tx1} \\ \varepsilon_i \end{pmatrix} \quad (13)$$

pour $i = 1, \dots, N$. Les vecteurs y_i , v_i , et ε_i sont de dimension $(T \times 1)$, x_i est de dimension $(KT \times 1)$, et W $(T \times L)$. Soient $\eta_i \equiv (y_i, v_i)'$, $\xi_i \equiv (x_i, \theta_i)'$ et $\zeta_i \equiv (0, \varepsilon_i)'$. Le système (13) s'écrit sous la forme d'un modèle structurel, reliant un vecteur de variables latentes endogènes η à un vecteur de variables latentes exogènes ξ , en terme d'analyse des covariances où le vecteur des effets individuels θ_i est traité comme un vecteur des variables latentes inobservées. Ceci donne pour $i = 1, \dots, N$:

$$\eta_i = \beta \eta_i + \Gamma \xi_i + \zeta_i \quad (14)$$

Il reste à spécifier les relations entre les variables latentes et les variables observées. Ces relations appelées équations de mesure pour η et ξ sont définies par $y = \Lambda_y \eta$ et $x = \Lambda_x \xi$. Nous supposons qu'il n'y a pas d'erreur de mesure sur les variables car y_i et x_i sont directement observées ce qui réduit le nombre de paramètres à estimer et simplifie l'écriture et le problème d'identification de ce modèle.

La méthode d'estimation

La méthode consiste à estimer la matrice de covariance Σ des variables observées y et x ⁽⁴⁾. A partir du modèle structurel (14) et des équations de mesure, cette matrice s'écrit sous la forme suivante :

⁽⁴⁾ Le lecteur pourra consulter Bentler (1983) pour un tour d'horizon de l'analyse structurelle des covariances.

$$\Sigma = \begin{pmatrix} \Lambda_y(I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1}\Lambda_y' & \Lambda_y(I-B)^{-1}\Gamma\Phi\Lambda_x' \\ \Lambda_x\Phi\Gamma(I-B)^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_x' \end{pmatrix} \quad (15)$$

où $\Phi = E(\xi\xi')$, $\Psi = E(\zeta\zeta')$. Cette matrice est fonction d'un vecteur de paramètres $\lambda^{(5)}$. Celui-ci contient tous les paramètres inconnus des matrices B , Γ , Λ_y , Λ_x , Φ , Ψ . Les matrices B , Λ_y et Λ_x sont constituées d'éléments connus. Les paramètres inconnus se limitent au vecteur β dans la matrice Γ , à σ^2 dans Ψ et aux sous-matrices $\Sigma_{x,x}$, $\Phi_{x,\theta}$ et Δ de la matrice Φ . Ces dernières notations désignent respectivement la matrice de covariance de x , la matrice de covariance entre x et θ et la matrice de covariance de θ .

Dans le cadre de l'estimation de frontière de production, il est intéressant d'estimer les effets individuels θ_i , ce qui est possible par l'intermédiaire de l'estimation des variables latentes. Ces estimations, les *factor scores*, sont obtenues à partir d'une régression des variables η et ξ sur les variables observées et s'écrivent $(\eta\xi)' = \Omega\Sigma^{-1}(yx)'$ où Ω est une matrice dont l'expression dépend des matrices B , Γ , Λ_y , Λ_x et Φ .

L'analyse structurelle des covariances permet non seulement d'estimer les paramètres usuels mais aussi toutes les corrélations entre les variables exogènes x et les effets individuels θ ainsi que celles entre les termes θ . Elle permet aux données de révéler les corrélations entre les variables explicatives et les effets individuels sans aucune hypothèse a priori sur leur existence.

RÉSULTATS EMPIRIQUES

Cette section présente l'estimation de la frontière de production et de mesures d'efficacité technique à partir d'un échantillon de panel d'exploitations céréalières, à l'aide des trois méthodes décrites dans la section précédente.

Les données

Notre échantillon de données provient d'enquêtes annuelles effectuées auprès d'exploitations agricoles françaises. Ce fichier, appelé le RICA (Réseau d'information comptable agricole), contient des données comptables et diverses caractéristiques sur les exploitations. Nous avons

⁽⁵⁾ L'estimateur $\hat{\lambda}$ est obtenu en minimisant une fonction d'ajustement $F(S, \Sigma(\lambda))$ dont l'expression sous certaines hypothèses (en particulier, normalité des observations) est égale à $\log \det(\Sigma) + \text{tr}(S \Sigma^{-1}) - \log \det(S) - P$ où P est le nombre de variables observées et S est la matrice de covariance empirique. Le lecteur pourra consulter Ivaldi, Perrigne et Simioni (1994) pour des références à d'autres méthodes.

extrait un sous-échantillon d'exploitations à dominante céréalière appelé OTEX 11 de 1982 à 1986. Suite aux opérations de cylindrage, nous avons obtenu un échantillon de 81 exploitations observées sur 5 années⁽⁶⁾. La production de céréales représente en moyenne 78 % de la production totale de chaque exploitation. La taille moyenne de ces unités de production est d'environ 70 hectares. Quelques indicateurs calculés sur la période considérée ont montré que la production est devenue plus intensive et que la taille des exploitations mesurée par leur superficie s'est accrue. De telles évolutions ont été observées sur tout le secteur céréalier (voir Carles, 1991).

Nous avons utilisé la méthode des indices de Törnqvist (voir Dievert, 1976) pour agréger les données du RICA en un output et trois inputs : capital, travail et consommations intermédiaires. Ces dernières variables ont été calculées en adaptant les travaux de Monier et Ossard (1990) à la production céréalière.

Résultats obtenus avec les méthodes *Within* et *GLS*

Le modèle à estimer s'écrit en utilisant une fonction de production de type Cobb-Douglas :

$$\log y_{it} = \beta_0 + \beta_1 \log k_{it} + \beta_2 \log l_{it} + \beta_3 \log m_{it} - \alpha_i + \varepsilon_{it}, \quad (16)$$

pour $i = 1, \dots, 81$ et $t = 1, \dots, 5$, où y , k , l , et m sont les indices quantité du produit, du capital, du travail et des consommations intermédiaires. Nous posons $\gamma_i = \beta_0 - \alpha_i$, d'où :

$$\log y_{it} = \beta_1 \log k_{it} + \beta_2 \log l_{it} + \beta_3 \log m_{it} + \gamma_i + \varepsilon_{it}, \quad \forall i = 1, \dots, 81 \text{ et } t = 1, \dots, 5 \quad (17)$$

Le tableau 1 donne les résultats des estimations selon les deux méthodes où \overline{TE} désigne la valeur moyenne de l'efficacité technique. La méthode *Within* semble donner de mauvais résultats au vu des coefficients de la fonction de production et de leur faible significativité. L'estimateur *Within* ne semble pas adapté du fait de la faible variation intra-individuelle des données.

Les résultats obtenus par la méthode *GLS* semblent meilleurs car cette méthode est plus performante lorsque le nombre de périodes est peu élevé. La faible valeur de $\overline{R^2}$ montre néanmoins que cette méthode est mal adaptée aux données. En effet, le modèle à effets aléatoires suppose que les variables exogènes ne sont pas corrélées avec l'effet individuel. Or cette hy-

⁽⁶⁾ On peut faire l'hypothèse que cette opération de cylindrage ne provoque pas de biais sur l'échantillon obtenu. En effet, le RICA renouvelle de façon aléatoire et annuelle une partie de son échantillon. Le cylindrage est alors indépendant de la disparition ou de la création des exploitations.

pothèse ne semble pas réaliste dans l'estimation de frontière de production ainsi que nous l'avons discuté dans la section précédente.

Le test d'Hausman (1978) permet de tester cette hypothèse. Posons l'hypothèse $H_0 : E(\alpha_i | x_i) = 0$ contre l'hypothèse $H_1 : E(\alpha_i | x_i) \neq 0$. Sous l'hypothèse nulle l'estimateur β obtenu par la méthode *Within* converge vers celui obtenu par la méthode *GLS*. Sous l'hypothèse alternative, l'estimateur *GLS* de β est biaisé. Nous allons comparer ces deux estimateurs.

Soit $\hat{q} = \hat{\beta}_W - \hat{\beta}_{GLS}$, la statistique de test s'écrit $m = \hat{q}' V(\hat{q})^{-1} \hat{q}$ où $V(\hat{q})$ est la variance de \hat{q} ⁽⁷⁾. Sous H_0 , la valeur m est distribuée selon une loi de χ^2 à K degrés de liberté. Nous obtenons après calculs une valeur de l'ordre de 421. Nous rejetons donc l'hypothèse d'indépendance entre x et α_i , c'est-à-dire que nous pouvons en déduire la présence de corrélation entre les trois inputs et l'efficacité individuelle. Nous n'avons aucune information a priori qui nous permette d'affirmer lequel de ces trois facteurs est effectivement corrélé avec l'effet individuel⁽⁸⁾.

Tableau 1.
Résultats des
estimations

Paramètres	β_0	β_1	β_2	β_3	σ^2	$\overline{R^2}$	\overline{TE}
<i>Within</i>	1,511	0,034 (1,202)	-0,003 (-0,079)	0,302 (4,328)	0,184 (28,46)	0,895	48 %
<i>QGLS</i>	0,535	0,081 (3,44)	0,061 (1,687)	0,646 (15,024)	0,198 (28,46)	0,495	66 %

Les nombres entre parenthèses sont les t de Student

Résultats obtenus avec l'analyse structurelle
des covariances

Nous proposons ici d'estimer une frontière stochastique de production de la forme suivante avec un effet individuel fonction linéaire du temps⁽⁹⁾:

$$\begin{cases} \log y_{it} = \beta_0 + \beta_1 \log k_{it} + \beta_2 \log l_{it} + \beta_3 \ln m_{it} + v_{it} \\ v_{it} = \theta_{0i} + \theta_{1i} t + \varepsilon_{it} \end{cases} \quad (18)$$

⁽⁷⁾ Cette variance s'écrit comme la différence des deux variances des deux estimateurs $\hat{\beta}_W$ et $\hat{\beta}_{GLS}$.

⁽⁸⁾ Un test d'Hausman pour chaque β est possible mais un test séparé perd toute sa puissance. De surcroît, si on effectue le test pour un des β , disons β_1 , ce test suppose que les variables associées aux paramètres β_2 et β_3 ne sont pas corrélées avec l'effet individuel. Il est évident que si des tests séparés donnent comme résultat plus d'un β qui rejettent l'hypothèse d'indépendance, ceci ne peut nous donner qu'une indication sur les variables corrélées avec l'effet individuel. Le test perd ainsi de sa valeur. Il est intéressant de noter que des tests séparés conduisent ici à rejeter H_0 pour β_2 et β_3 , résultat qui ne se confirmera qu'en partie par la suite.

⁽⁹⁾ L'introduction d'un terme quadratique ne fournissait pas de meilleurs résultats.

Ce modèle général peut être estimé selon différentes hypothèses⁽¹⁰⁾. Nous acceptons l'hypothèse de fixité dans le temps des covariances entre x et θ . Nous acceptons également l'hypothèse d'efficacité individuelle variant dans le temps et posons les covariances entre les inputs et les effets individuels égales à zéro à l'exception de celle entre la quantité de consommations intermédiaires et θ_1 qui est significativement différente de zéro. Ceci réduit le nombre des paramètres du modèle à 8. Les résultats sont donnés dans le tableau 2 où les variables δ_0 , δ_1 , $\delta_{0,1}$ dénotent les estimations des éléments de la matrice Δ .

Tableau 2.
Résultats des
estimations

β_1	β_2	β_3	σ_{m,θ_1}	δ_0	δ_1	$\delta_{0,1}$	σ^2
0,042	0,108	0,745	-0,005	0,052	0,002	-0,005	0,017
(2,18)	(3,47)	(17,89)	(-4,51)	(4,58)	(3,56)	(-2,32)	(10,95)

Les nombres entre parenthèses correspondent aux t de Student

Le tableau 3 donne quelques statistiques sur les estimations des deux valeurs θ_{0i} et θ_{1i} . Nous pouvons calculer les effets individuels variant dans le temps et les exprimer sous la forme $\hat{v}_{it} = \hat{\theta}_{0i} + \hat{\theta}_{1i} t$. Nous pouvons alors en déduire des mesures individuelles d'efficacité variant dans le temps selon la formulation de Cornwell, Schmidt et Sickles (1990).

$$TE_{it} = 100 \times \varepsilon^{-\hat{\alpha}_{it}} \text{ avec } \hat{\alpha}_{it} = \max_{j \neq i} (\hat{v}_{jt}) - \hat{v}_{it}.$$

Le tableau 3 donne également quelques statistiques sommaires sur les distributions annuelles de ces mesures individuelles d'efficacité. Nous pouvons en conclure qu'il y a en moyenne une nette décroissance des niveaux d'efficacité dans la période considérée. De plus, l'écart entre l'exploitation la plus efficace et la moins efficace s'accroît sur la période.

Tableau 3.
Statistiques relatives
aux mesures
d'efficacité

		Moyenne	Ecart-type	Minimum	Maximum
θ_0		0,620	0,204	0,104	1,075
θ_1		0,079	0,034	-0,078	0,175
TE	(1982)	68,17	13,50	39,79	100,00
TE	(1983)	65,39	12,89	38,06	100,00
TE	(1984)	61,22	12,35	32,21	100,00
TE	(1985)	57,37	12,13	25,75	100,00
TE	(1986)	53,82	12,14	25,59	100,00

⁽¹⁰⁾ Des hypothèses peuvent être testées en estimant les différents modèles contraints correspondants. Une recherche de spécification entre différents modèles emboîtés peut s'effectuer en comparant deux à deux les mesures de χ^2 de chacun des modèles. Voir Ivaldi, Perrigne, Simioni (1994) pour une discussion de ces tests.

L'utilisation d'une méthode basée sur l'analyse structurelle des covariances a permis de révéler d'une part, l'existence d'une corrélation négative entre les consommations intermédiaires et le coefficient du temps de l'effet individuel et d'autre part, la variabilité temporelle des efficacités techniques des exploitations. L'étendue des valeurs des paramètres θ_0 et θ_1 montre l'hétérogénéité des exploitations céréalières en termes d'efficacité productive. Le terme θ_1 allant de valeurs négatives à des valeurs positives montre les différentes situations des exploitations en termes d'évolution. Ces deux observations pourraient faire l'objet d'une étude qui permettrait de caractériser cette hétérogénéité et ces évolutions et ainsi de les expliquer. L'étendue des valeurs pour θ_1 expliquerait l'évolution en moyenne décroissante des niveaux d'efficacité car d'une part, certains individus (les moins efficaces) ont une efficacité décroissante ($\theta_1 < 0$) et d'autre part, l'écart entre les exploitations les plus performantes et les moins performantes s'accroît sur la période considérée. Les consommations intermédiaires étant principalement constituées d'engrais, la corrélation négative entre ces dernières et le terme pouvant s'interpréter comme une faculté individuelle d'intégration du progrès technique nous amène à penser qu'il y aurait eu une tendance à la sur-utilisation de cet input par les exploitations et qu'une réduction de son utilisation a contribué à améliorer leur productivité. La justification économique de ce résultat dû probablement à l'accroissement des prix des engrais sur la période considérée (voir Carles, 1991) reste donc une question qui mériterait d'être développée en termes d'inefficacité allocative plus que technique.

CONCLUSION

Nous avons commenté dans ce papier diverses méthodes d'estimation de frontière de production et des niveaux individuels d'efficacité dans le cadre de données de panel en essayant de tenir compte des spécificités d'un échantillon d'exploitations agricoles. Les résultats obtenus avec la méthode *Within* ont montré la limite de celle-ci lorsque l'échantillon dont on dispose contient peu de périodes. L'estimation de la frontière par la méthode des moindres carrés généralisés a permis de mettre en lumière l'existence d'une éventuelle corrélation entre les variables expliquées et l'effet individuel représentant l'efficacité.

L'utilisation d'une méthode plus complexe basée sur l'analyse structurelle des covariances a permis de révéler d'une part, l'existence d'une corrélation négative entre la quantité utilisée de consommations intermédiaires et l'effet individuel, et d'autre part, des mesures d'efficacité variant dans le temps. Les résultats empiriques de cette méthode ont suggéré divers éléments qui ont permis de mieux comprendre l'évolution des exploitations en terme d'efficacité productive par rapport à ce qu'une approche plus classique aurait pu apporter. Il reste néanmoins de

nombreux points qui demanderaient une recherche plus approfondie, comme l'amélioration de la productivité des exploitations liée à une diminution de la quantité utilisée de consommations intermédiaires, ou encore la justification de l'évolution décroissante des niveaux d'efficacité dans un contexte d'une production de plus en plus intensive et de plus en plus spécialisée.

Le premier résultat pourrait être lié à un problème d'efficacité allocative, donc d'allocation du facteur en fonction de son prix, puisqu'il a été observé un accroissement du prix des engrais sur la période considérée. L'étude de ce problème pourrait être complétée par l'estimation de fonction de coût et de mesures d'efficacité allocative. Quant au second résultat, la décroissance des niveaux d'efficacité doit être relativisée de par leur calcul. En effet, le meilleur individu restant au même niveau 100% d'efficacité, l'écart entre le meilleur et les autres aurait tendance à croître au cours du temps. Si on regarde l'estimation du paramètre θ_i (voir tableau 3), il est positif à l'exception de quelques exploitations, ce qui signifie que l'effet individuel est en général une fonction croissante du temps. De plus, il semble qu'il y ait dans l'échantillon une grande différence entre les exploitations efficaces, situées principalement dans les grandes plaines céréalières du Nord de la France pour la culture du blé et dans les Landes pour la culture du maïs, et les autres nettement moins efficaces. Une étude ultérieure pourrait viser à la compréhension de ces différences entre exploitations.

Il est à noter que les résultats économiques suggérés par l'étude concernent une sous-population d'exploitations que couvre le RICA. Ce dernier considérant des exploitations d'assez grande taille, les résultats ne peuvent pas donc être généralisés à l'ensemble des exploitations céréalières. Enfin, il serait intéressant de disposer de cohortes d'exploitations sur une plus longue période pour lier l'étude de l'efficacité productive à la disparition des exploitations.

BIBLIOGRAPHIE

- AIGNER (D.J.), LOVELL (C.A.K.) et SCHMIDT (P.), 1977 — Formulation and estimation of stochastic frontier production models, *Journal of Econometrics*, 6, pp. 21-37.
- BATTESE (G.M.), 1992 — Frontier production functions and technical efficiency: a survey of empirical application in agricultural economics, *Agricultural Economics*, 7, pp. 185-208.

- BENTLER (P.M.), 1983 — Simultaneous equation system as moment structure models. With an introduction to latent variable models, *Journal of Econometrics*, 22, pp.13-42.
- CARLES (R.), 1991 — Les perspectives des exploitations céréalières, *Cahiers d'Economie et Sociologie Rurales*, 18-19, pp. 136-158.
- CORNWELL (C.), SCHMIDT (P.) et SICKLES (R.), (1990) — Production frontiers with cross-sectional and time-series variations in efficiency levels, *Journal of Econometrics*, 46, pp. 185-200.
- DIEWERT (W.E.), 1976 — Exact and superlative index numbers, *Journal of Econometrics*, 4, pp. 15-145.
- HAUSMAN (J.A.), 1978 — Specification test in econometrics, *Econometrica*, 46, pp. 1251-1271.
- HAUSMAN (J.A.) et TAYLOR (W.E.), 1981 — Panel data and unobservable individual effects, *Econometrica*, 49, pp. 1377-1398.
- HSIAO (C.), 1986 — *Analysis of panel data*, Econometric Society monographs, Cambridge, Cambridge University Press.
- IVALDI (M.), PERRIGNE (I.) and SIMIONI (M.), 1994 — Productive efficiency of French grain producers: a latent variable model, *Journal of Productivity Analysis*, 5, pp. 287-299.
- JÖRESKOG (K.G.), 1970 — A general method for analysis of covariance structures, *Biometrika*, 57, pp. 239-251. Reprinted as chapter 12 in: AIGNER (D.J.), and GOLDBERGER (A.S) (eds.), 1977, *Latent Variables in Socio-Economic Models*, Amsterdam, North-Holland.
- JÖRESKOG (K.G.), 1977 — Structural equation models in the social sciences: specification, estimation, testing, in: KRISHNAIAH (P.R.) (ed.), *Applications of Statistics*, Amsterdam, North-Holland, pp. 265-287.
- MONIER-DILHAN (S.) et OSSARD (H.), 1990 — Méthode de construction d'une base de données de panel pour l'économie de la production à partir d'un échantillon d'exploitations du RICA, Document de travail INRA, Toulouse.
- PERRIGNE (I.), 1991 — Frontière de production et mesure d'efficacité technique: application aux exploitations céréalières, Thèse, Université de Toulouse I, 179 p.
- SCHMIDT (P.), 1985-86 — Frontier production functions, *Econometric Reviews*, 4, pp. 289-328.
- SCHMIDT (P.) and SICKLES (R.), 1984 — Production frontiers and panel data, *Journal of Business and Economic Statistics*, 4, pp. 367-374.