Accounting for unobserved heterogeneity in micro-econometric agricultural production models: a random parameter approach

Alain Carpentier[1.2.]
Fabienne Femenia[1.]
Philippe Koutchade[1.]

1. INRA UMR 1302 SMART
4. Allée Adolphe Bobierre
35011 Rennes
France
alain.carpentier@rennes.inra.fr

2. Agrocampus Ouest UMR 1302 SMART
4. Allée Adolphe Bobierre
35011 Rennes
France

Selected Poster prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meeting. Minneapolis. MN. July 27-29. 2014

ACCOUNTING FOR UNOBSERVED HETEROGENEITY IN MICRO-ECONOMETRIC AGRICULTURAL PRODUCTION MODELS: A RANDOM PARAMETER APPROACH

ALAIN CARPENTIER*°, FABIENNE FEMENIA*, PHILIPPE KOUTCHADE*,

*INRA, UMR 1302 SMART, 4 ALLÉE ADOLPHE BOBIERRE, 35011 RENNES, FRANCE

°AGROCAMPUS OUEST , UMR 1302 SMART, 4 ALLÉE ADOLPHE BOBIERRE, 35011 RENNES, FRANCE

MOTIVATIONS

Farmers face different production conditions due to hetegeneous soil quality or usual climatic conditions accross space. They also own different machineries and different wealth levels. Finally, farmers are also different because of their various educational level or abilities, as well as because they may have different objectives with respect to income risk or with respect to the leisure *versus* labour trade-off. These heterogeneity sources are likely to have important impacts on farmers' production choices. But to control for the effects of these heterogeneity sources is difficult . First, many heterogeneity sources are not suitably described in the data sets usually used by agricultural economists. Second, these heterogeneity sources are numerous. As a result, empirical investigators are usually forced to focus on a few of these heterogeneity sources when specifying their empirical models.
In this study we take for granted that farmers' choices rely on heterogeneous determinants by considering random parameter models. *I.e.*, we account for heterogeneity in farmers' production choice models not by considering that the model relevant to a given farmer relies on a parameter vector specific to this farmer. Of course, the use of control variables is allowed in our modelling approach. But it is omitted for simplicity as well as for investigating the potential of random parameter models. Such models are already used in other economic litteratures, *e.g.* in labour economics (see, *e.g.*, Bonhomme and Robin, 2009) or in empirical industrial organization (Ackerberg et al, 2007). Random parameter models are now routinely used by empirical investigators considering economic discrete choices (Train, 2009).

MODELLING FRAMEWORK

We consider short run production choices of farmers, *i.e.* an acreage (share) demand system and yield supply system.
We use panel data so that observations are indexed by i (farm/farmers) and t (year). A (parametric) random parameter model is composed of two parts.
The first part of the model, the core model, describes the process of interest. The equation $\mathbf{x}_{it} = \mathbf{g}(\mathbf{z}_{it};\mathbf{a}_0,\mathbf{q}_i;\mathbf{e}_{it})$ decribes the choices $\mathbf{x}_{it}$ of farmer i in year t as a response function g to the observed determinants $\mathbf{z}_{it}$ and to the unobserved determinants $\mathbf{e}_{it}$ (error terms) defined according to the farmer specific parameter vector $\mathbf{q}_i$. The term $\mathbf{a}_0$ is a fixed paramter vector. The core model is parametric, as it is the case here, if the probability distribution of $\mathbf{e}_{it}$ conditional on $\mathbf{z}_{it}$ and $\mathbf{q}_i$ is given. The second part of the model describes the parametric model, $\mathbf{q}_i \sim \mathcal{D}(\mathbf{b}_0)$, of the probability distribution of the random parameters $\mathbf{q}_i$. This probability distribution describes the distribution of the $\mathbf{b}_i$ across the considered famers' population. The more the $\mathbf{q}_i$ are variable across farmers, the more heterogeneity matters for modelling farmers' choices. The parameters to be estimated are $\mathbf{a}_0$, the fixed parameter vector of the core model, and $\mathbf{b}_0$, the parameter vector of the probability distribution of $\mathbf{q}_i$.
The core model of a random parameter models accounting for farms and farmers' heterogeneity can be interpreted as a local approximation of what might be the « true » model of each farmer's choice process. Of course, the core model needs to be flexible enough for adequately representing farmers' choices. Similarly, the choice of the probability distribution of the random parameters plays a crucial role for the abilility of the specified statistical mixture model to suitably account for the heterogeneity of the farmers decision process in the considered sample.

DATA AND ESTIMATION APPROACH

We use a panel data set of 111 French grain producers observed from 2004 to 2007 obtained from the European Farm Accountancy Data Network (FADN). The considered farmers are located in northern France. The crop price indices are computed from the observed data while the variable input price index is computed by considering the observed input uses (fertilisers, pesticides, energy and seeds) and price indices obtained from the French Agricultural Department. We do not model variable input uses in this preliminary study. To extend our modelling framework is relatively involved with input uses observed at the farm level such as those provided by the French FADN.
Whereas econometricians mainly rely on Simulated Maximum Likelihood (ML) estimators for estimating the parameters of random parameter models, statisticians usually employ ML estimators computed by Stochastic Expectation-Maximisation (SEM) algorithms (McLachlan and Krishnan, 2008). In this study we compute ML estimators of the model parameters by using the SEM algorithm proposed by Delyon et al (1999).

EMPIRICAL MODEL

The model considered in this study can be interpreted as a random parameter version of the model considered by Carpentier and Letort (2013) . This model combines a Nested MNL acreage share model with quadratic yield functions. It assumes that farmers maximize their expected profit in two steps. First they maximize the expected return to each crop under the assumption that this return doesn't depend on the crop acreages. Second, farmers allocate land to the different crops for maximizing their expected profit provided that they incur implicit acreage management costs. These management costs provide incentive for crop diversification. In the 3 crop case with crop 3 (oilseeds) as the reference crop and crops 1 (wheat) and 2 (othser cereals) grouped into a nest, the core production choice model is composed of two equation systems, the acreage share system :

$$\begin{cases} s_{k,it} = \dfrac{\exp(\rho_i \pi_{k,it})}{\sum_{k=1,2}\exp(\rho_i \pi_{k,it})} \dfrac{\exp\left(\alpha_i \rho_i^{-1} \ln \sum_{k=1,2}\exp(\rho_i \pi_{k,it})\right)}{\exp(\rho_i \pi_{3,it}) + \exp\left(\alpha_i \rho_i^{-1} \ln \sum_{k=1,2}\exp(\rho_i \pi_{k,it})\right)} \quad \text{for } k=1,2 \\ \text{with } \begin{cases} \pi_{k,it} = p_{k,it}\beta_{k,i} + 1/2 \times a_{k,0} w_{it}^2 p_{k,it}^{-1} - d_{k,i} + u_{k,it} \quad \text{for } k=1,2 \\ \pi_{3,it} = p_{3,it}\beta_{3,i} + 1/2 \times a_{3,0} w_{it}^2 p_{3,it}^{-1} \end{cases} \end{cases}$$

with $s_{3,it} = 1 - s_{1,it} - s_{2,it}$, and the yield supply system:

$$\left\{ y_{k,it} = \beta_{k,i} - 1/2 \times a_{k,0} w_{it}^2 p_{k,it}^{-2} + v_{k,it} \quad \text{for } k=1,2,3. \right.$$

It is assumed that the error terms $\mathbf{u}_{it}$ and $\mathbf{v}_{it}$ are *i.i.d.* across years and farms and that they are independent from each other. We basically assume that the dynamic features of the agricultural production and choices processes are sufficiently stable for being suitably accounted for by the random parameter specification. The terms $\mathbf{u}_{it}$ and $\mathbf{v}_{it}$ can be assumed to be independent because $\mathbf{u}_{it}$ is unknown to farmer i when choosing $\mathbf{s}_{it}$.
The parametric random parameter model is completed by distribution assumptions related to the error terms of the core model and to the random parameter vector:

$$\mathbf{u}_{it} \sim \mathcal{N}(\mathbf{0},\mathbf{\Psi}_0),\ \mathbf{v}_{it} \sim \mathcal{N}(\mathbf{0},\mathbf{\Lambda}_0) \text{ and } \mathbf{q}_i \sim \mathcal{N}(\boldsymbol{\mu}_0,\mathbf{\Omega}_0) \text{ for } t=1,\dots,T \text{ and } i=1,\dots,N.$$

The error terms $(\mathbf{u}_{it},\mathbf{v}_{it})$, the explanatory variables $\mathbf{z}_{it}$ and the random parameters $\mathbf{q}_i$ are assumed to be mutually independent.

Observed variables $\mathbf{x}_{it} \equiv (\mathbf{s}_{it},\mathbf{y}_{it})$ and $\mathbf{z}_{it} \equiv (\mathbf{p}_{it}, w_{it})$

$s_{k,it}$: acreage share of crop k

$y_{k,it}$: yield of crop k

$p_{k,it}$: price index of output k

w_{it} : price index of the variable input aggregate

Error terms $\mathbf{e}_{it} \equiv (\mathbf{u}_{it},\mathbf{v}_{it})$

$u_{k,it}$: random part of the yield of crop k

$v_{k,it}$: random part of the per hectare management cost of crop k $(v_{3,it} \equiv 0)$

Random parameters $\mathbf{q}_i \equiv (\ln\boldsymbol{\beta}_i, \ln\alpha_i, \ln\rho_i, \mathbf{d}_i)$ and $\mathbf{b}_0 \equiv (\boldsymbol{\mu}_0, \mathbf{\Psi}_0, \mathbf{\Lambda}_0, \mathbf{\Omega}_0)$

$\beta_{k,i}$: yield mean level of crop k for farmer i

α_i : acreage cost function flexibility parameter across the "cereal" and "oilseed and protein crop" nests

ρ_i : acreage cost function flexibility parameter within the "cereal" nest

$d_{k,i}$: per hectare management cost of crop k $(d_{3,i} \equiv 0)$

CONCLUSION: HETEROGENEITY MATTERS

Estimation results show that the model fits relatively well to the data. Most parameters -- especially the fixed parameter $\mathbf{a}_0$, the expectation of the random parameters $\boldsymbol{\beta}_i$ and $\mathbf{d}_i$, and the variance matrices of the error terms $\mathbf{u}_{it}$ and $\mathbf{v}_{it}$ – appear to be precisely estimated. As will be shown below, some random parameter covariances appear to be less precisely estimated. This might be explained by the limited size of the considered sample.
The yield equation parameters are precisely estimated. This was expected since each yield equation basically is a regression equation with individual random terms. The parameter estimates lie in reasonable ranges. In particular, the price effect parameter parameter $\mathbf{a}_0$ is positive. The estimates of the probability distribution of $\boldsymbol{\beta}_i$ show that the $\beta_{k,i}$ parameters significantly vary accross farms while being strongly positively correlated to each other. This was expected because the yield potential vary accross regions, as well as because good growing conditions for a grain crop are also good for the others. The variance of $\beta_{k,i}$ is equal to that of $u_{k,it}$ for cereals, but the variance of $u_{k,it}$ is four times that of in the oilseeds case. Provided that rapesseed is by far the most imporant oilseed in northern France, this may reflect the fact that the rapeseed yield is more risky than the cereal yield, mostly due to bugs and diseases.
The acreage share equation parameter estimates also range in reasonable ranges. The estimated expectation of α_i, respectively of ρ_i, equals 0.149, respectively 0.223. These expectation estimates are higher to the corresponding fixed parameter estimates obtained by Carpentier and Letort (2013) with similar, albeit different, models and data. Importantly, the estimate of the expectation of ρ_i is higher than that of α_i. This is a sufficient condition for the entropic acreage management cost function lying at the root of the Nested MNL acreage share function to be convex. According to the estimates of their respective variances, the α_i and ρ_i parameters significantly vary accross farms. This result is important for simulation studies because these parameters largely determine the acreage price elasticities in MNL acreage share models. The higher α_i and ρ_i are, the more reactive the acreages are to price changes. As a matter of fact the parametric considered in this study allows computation of estimates of α_i and ρ_i for each farm of the sample, according to the logic 'tell me what you do, I'll tell you who you are'. The parameter estimates obtained here can be used to compute an estimate of the expectation α_i (indeed of any element of $\mathbf{q}_i$) conditionnal on $(\mathbf{y}_{it}, \mathbf{s}_{it}, \mathbf{z}_{it})$ for $t = 1,\dots,T$. Although the corresponding parameters estimates are not very precise, the elements of $\boldsymbol{\beta}_i$ appear to be positively correlated with α_i. A possible interpretation of this result is as follows. High levels of $\boldsymbol{\beta}_i$ indicate good farming conditions for grain crops in farm i and/or farmer i technical ability. This implies that the farm operation is sufficiently profitable for allowing suitable machinery investments which, in turn, implies a high level of α_i and, finally, relatively unconstrained acreage choices.

Our results tend to show that heterogeinity matters in agricultural production choice models. Estimates of random parameter models such as the one presented here can be used for, at least, two purposes: for the calibration of simulation models accounting for farm unobserved heterogeneity and for investigating potential explanations of this unobserved heterogeneity.

Table 1. Selected parameter estimates, yield equations

	$a_{k,0}$	$E(\ln\beta_{k,i})$	$Cov(\ln\beta_{k,it},\ln\beta_{\ell,it})$		
			Wheat $(\ell=1)$	**Oth. cereals** $(\ell=2)$	**Oilseeds** $(\ell=3)$
Wheat $(k=1)$	**1.579***	**2.175***	**0.012***	**0.012***	**0.008***
Oth. cereals $(k=2)$	**0.989***	**2.120***	**0.012***	**0.017***	**0.011***
Oilseeds $(k=3)$	**0.862***	**1.804***	**0.008***	**0.011***	**0.008***

Table 2. Selected parameter estimates, acrege share equations

	Expectation	**Covariance with**				
		$\ln\alpha_i$	$\ln\rho_i$	$\ln\beta_{1,i}$	$\ln\beta_{2,i}$	$\ln\beta_{3,i}$
				Wheat	**Oth. cereals**	**Oilseeds**
$\ln\alpha_i$	**-2.327***	**0.136***	**0.061**	**0.006**	**0.012**	**0.008**
$\ln\rho_i$	**-2.039***	**0.061**	**0.188***	**-0.016**	**-0.011**	**-0.006**

* indicates rejection of the null hypothis at the 5% level

REFERENCES

Ackerberg, D., C.L. Benkard, S.T. Berry and A. Pakes. 2007. Econometric tools for analyzing market outcomes. The Handbook of Econometrics. Heckman, J.J. and E. Leamer (eds). Chapter 63. North-Holland ; **Bonhomme, S. and J.-M. Robin. 2009.** Assessing the equalizing force of mobility using short panels: France, 1990-2000. *Review of Economic Studies*, 76(1):63-92 ; **Carpentier, A. and E. Letort. 2013.** Multicrop production models with Multinomial Logit acreage shares. *Environmental and Resource Economics*, forthcoming ; **Delyon, B., M. Lavielle and E. Moulines. 1999.** Convergence of a stochastic approximation of the EM algorithm. *Annals of Statistics*, 27(1):94-128 ; **McLachlan, G. and T. Krishnan. 2008.** *The EM algorithm and extensions*, 2nd edition. Wiley Edition ; **Train, K. 2009.** *Discrete choice models with simulation*, 2nd edition. Cambridge Unuversity Press.

Alain Carpentier, Senior Researcher Email: a.carpentier@rennes.inra.fr
Fabienne Femenia, Research Fellow, Email: f.femenia@rennes.inra.fr
Philippe Koutchade, PhD, Email: p.koutchade@rennes.inra.fr