Modeling heterogeneous technologies in the presence of sample selection: The case of dairy farms and the adoption of agri#environmental schemes in France

Journal Article

Author(s):
Dakpo, K. Hervé; Latruffe, Laure; Desjeux, Yann; Jeanneaux, Philippe

Publication date:
2022-05

Permanent link:
https://doi.org/10.3929/ethz-b-000519526

Originally published in:
Agricultural Economics 53(3), https://doi.org/10.1111/agec.12683

Received: 22 June 2020 | Revised: 30 March 2021 | Accepted: 10 October 2021

DOI: 10.1111/agec.12683

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists
IAAE
WILEY

ORIGINAL ARTICLE

Modeling heterogeneous technologies in the presence of sample selection: The case of dairy farms and the adoption of agri-environmental schemes in France

K Hervé Dakpo[1,2] | Laure Latruffe[3] | Yann Desjeux[3] | Philippe Jeanneaux[4]

[1] INRAE, AgroParisTech, Economie Publique, Université Paris-Saclay, Thiverval-Grignon, France

[2] Agricultural Economics and Policy Group, ETH Zürich, Zürich, Switzerland

[3] INRAE, GREThA, Université de Bordeaux, Pessac, France

[4] UMR Territoires, VetAgro Sup, Lempdes, France

Correspondence
K Hervé Dakpo, Université Paris-Saclay, INRAE, AgroParisTech, Economie Publique, 78850, Thiverval-Grignon, France.
Email: k-herve.dakpo@inrae.fr

Data Appendix Available Online: A data appendix to replicate the main results is available in the online version of this article. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Abstract

In this article, we assess farms' technical efficiency accounting for their production heterogeneity and correcting for the potential endogeneity associated with the adoption of Common Agricultural Policy (CAP) agri-environmental schemes (AESs). We estimate a first-step selection probit model. In a second step we estimate the latent class stochastic frontier model (LCSFM) separately on each of the two sub-samples, the AESs adopters and the AESs non-adopters. We also account for heteroscedasticity in the estimation of inefficiency effects in the LCSFM within each sub-sample. The application is to Farm Accountancy Data Network (FADN) dairy farms in France during 2002–2016. We identify one class with intensive technology and one class with extensive technology for each of the two sub-samples. The investigation of inefficiency effects shows that modeling production heterogeneity could help better target the CAP, since the relationship between operational subsidies and farms' efficiency differs depending on whether or not production heterogeneity is accounted for.

KEYWORDS

agri-environmental schemes, efficiency, farms, France, heterogeneous technologies, latent class, sample selection

JEL CLASSIFICATION

C01, D22, Q10, Q50

1 | INTRODUCTION

Evaluating the levels and determinants of farms' technical efficiency is a topic that has received considerable attention in the literature, using various benchmarking methods. One crucial shortcoming in benchmarking studies is that technological heterogeneity across firms is generally ignored. Not accounting for heterogeneity in the technology is a strong assumption and may lead to misspecification in the empirical evaluations (Hayami & Ruttan, 1970). In agriculture, the technologies are often complex, and their efficient use may be strongly affected by farmers' abilities and the conditions in which the farms operate.

Heterogeneous production technologies can be modeled through the meta-frontier framework. In this, a single characteristic is considered to define a priori homogeneous groups of observations (in terms of location, main production, organic or conventional agriculture, etc.) for which different technologies are estimated (Battese & Rao, 2002; Battese et al., 2004). However, from a methodological point of view, the meta-frontier approach may be affected by the problem of incompleteness, in the sense that groups of observations may not be correctly identified and some unobservable factors may not be considered. In addition, the meta-frontier estimation is, in fact, a three-stage approach in which, in the first stage, the sample is split into several groups, in the second stage, a frontier is estimated for each of these a priori defined groups, and in the third stage, the meta-frontier is constructed. These distinctive (independent) stages imply a loss of information about the common features of observations in different groups since each group frontier is constructed separately (Alvarez & del Corral, 2010).

An appealing method that has been proposed to circumvent the shortcomings of the meta-frontier is the technique of latent class modeling, coupled with the stochastic frontier framework (W. Greene, 2005; Orea & Kumbhakar, 2004). This technique simultaneously estimates the groups of farms, called the classes, and each class's frontier.

In agriculture, latent class modeling has been used by Alvarez and Arias (2015), Alvarez and del Corral (2010), Alvarez et al. (2012), Cillero et al. (2019), Kellermann and Salhofer (2014), Orea et al. (2015), Sauer and Paul (2013), and Grovermann et al. (2021). In general, farms are separated into classes depending on the degree of intensification of their practices, based, for example, on livestock density. Another aspect that may differentiate farms' technologies is whether farms have adopted agri-environmental schemes (AESs). AESs are part of the European Union's (EU) Common Agricultural Policy (CAP) and aim to increase farmers' adoption of environmentally-friendly practices. Farmers voluntarily adopt AESs, generally for 5 years, and receive payments to compensate for additional costs and potential profit losses following the adoption of environmental practices.

Many diverse AESs are designed depending on the EU Member States or the main production specialization of farms to accommodate production systems' heterogeneity (Primdahl et al., 2010). For instance, in France in the case of livestock farming, environmental practices covered by AESs relate to, among other things: the extent of permanent grassland; number of animals per hectare (stocking rate); no-tillage and no pesticides on permanent grassland; set-up of ecological interest area; grass buffer strips; low use of nitrogen fertilizers; and conversion to and maintenance of organic farming. Additionally, the AESs are implemented at the regional level in France, which introduces more disparity among the adopters (Chabé-Ferret & Subervie, 2013). This heterogeneity is likely to be reflected in the production technologies of farmers.

Modeling adoption of AESs as a variable that enables the separation of farms into classes in a stochastic frontier approach is not straightforward due to endogeneity. The literature on AES adoption highlights that several reasons might explain why a farmer chooses one practice over another, such as cost-benefit analysis, opportunistic behavior, etc. (Barreiro-Hurlé et al., 2010; Espinosa-Goded et al., 2013; Vanslembrouck et al., 2002). AES adoption may also depend on farmers' willingness and capacity to adjust their technology and practices. As such, the adoption process is endogenous and has to be explicitly accounted for in estimating production technologies.

In a classic efficiency analysis, we could use the stochastic frontier (SF) framework (Aigner et al., 1977) which has been extended to the Heckman selection models in W. Greene (2010) and Lai (2015). A few studies have examined the effect of technology choice on technical efficiency (Bostian et al., 2019; Bravo-Ureta et al., 2012; Henningsen et al., 2015; Kumbhakar et al., 2009; Mayen et al., 2010; Villano et al., 2015). However, none of them considered heterogeneity within the framework of latent class modeling.

Therefore, our aim in this article is to examine the question of farms' technological heterogeneity, taking into account that some farmers adopt AESs and others do not. From a methodological point of view, our article will contribute to the literature in two ways. We extend W. Greene (2010)'s SF sample selection model to account for production heterogeneity under the latent class stochastic frontier model (LCSFM) (Orea & Kumbhakar, 2004). In this framework, endogeneity arises from the correlation between the two-sided error component in the production function and the sample selection equation noise.[1] Our new model accommodates the two strands of the literature and considers production heterogeneity under the endogenous adoption of AESs. It further accounts for the inclusion of efficiency drivers (as heteroscedasticity in the one-sided error term). Empirically, we operationalize the new model by considering a sample of French dairy farms during 2002–2016. Finally, in the implementation of the LCSFM, while the prior probability of farms belonging to a class is generally fixed over time, in our new model we adopt a pooled version of the LCSFM with modified prior probabilities that change over specific periods (2002-2006, 2007—2013, and 2014–2016) which correspond to the main reforms of the CAP.

The rest of the article is structured as follows. Section 2 presents the selection correction for the LCSFM with heteroscedasticity, and Section 3 describes the database and the empirical specification. Section 4 explains the results, while Section 5 concludes.

2 | THE STOCHASTIC FRONTIER MODEL WITH SELECTION CORRECTION

The sample selection model initially introduced by Heckman (1976) and (1979) has been adapted to the case of SF by W. Greene (2010). Here we extend this model to account for technological heterogeneity under the latent class structure discussed in Orea and Kumbhakar (2004). First we present the SF model corrected for selection that is formulated in W. Greene (2010). Then we elaborate on our extension of the sample selection correction to the case of LCSFM.

2.1 | Greene's (2010) sample selection correction

2.1.1 | Likelihood construction

The SF model was originally and independently proposed by Aigner et al. (1977) and Meeusen and Vandenbroeck (1977). It is a composed error structure and is written as follows:

$$\mathrm{y_{it}} = \beta' \mathrm{x_{it}} + \mathrm{v_{it}} - \mathrm{u_{it}} i = 1, \dots, Q; t = 1, \dots, T \quad (1)$$

where

$$\mathrm{u_{it}} = \sigma_{\mathrm{u}} \left|\mathrm{U_{it}}\right|, \mathrm{U_{it}} \sim N(0,1)$$

$$\mathrm{v_{it}} \sim N(0, \sigma_{\mathrm{v}}^2) \quad (2)$$

where y_{it} is the logarithmic output quantity of each farm i in time t; x_{it} is the vector of transformations of input quantities; β is the vector of parameters to be estimated; $N(.)$ indicates the normal distribution; and $v_{it}-u_{it}$ is the composed error, where v_{it} is the unrestricted statistical noise (random variations) with variance σ_v^2 and u_{it} represents the inefficiency term with scale parameter σ_u. σ_v^2 is parameterized as exp(c).

The formulation above is extended to account for heteroscedasticity in the one-sided error term u.[2] We develop the model based on the literature (Caudill & Ford, 1993; Caudill et al., 1995; Reifschneider & Stevenson, 1991). More specifically, we have

$$\mathrm{u_{it}} = \sigma_{\mathrm{uit}} \left|\mathrm{U_{it}}\right|, \mathrm{U_{it}} \sim N(0,1)$$

$$\sigma_{\mathrm{uit}}^2 = exp(\delta' \mathrm{H_{it}}) \quad (3)$$

where H_{it} is a vector of variables (including a constant term), namely the drivers of inefficiency; and δ is the vector of parameters to be estimated.

The estimation of model (1) can be conducted using various procedures. In the case of the sample selection, W. Greene (2010) advocated the use of the maximum simulated likelihood (MSL) for its practicability. However, here, we used the quadrature method, which is more efficient than the MSL.

Conditional on u_{it}, the density of the two-sided error disturbances v_{it} is:

$$\mathrm{f}(\mathrm{y_{it}}|\mathrm{x_{it}}, |\mathrm{U_{it}}|) = \frac{1}{\sigma_{\mathrm{v}}} \phi \left(\frac{\mathrm{y_{it}} - \beta' \mathrm{x_{it}} + \sigma_{\mathrm{uit}} |\mathrm{U_{it}}|}{\sigma_{\mathrm{v}}} \right) \quad (4)$$

where ϕ is the density function of the standard normal distribution.[3]

The log-likelihood function is obtained by integrating the density in (4) over the range of $|U_{it}|$. Thus:

$$\log \mathrm{f}\,(\mathrm{y_{it}}|\mathrm{x_{it}}) = \log \int_{\mathrm{U_{it}}} \frac{1}{\sigma_{\mathrm{v}}} \phi \left(\frac{\mathrm{y_{it}} - \beta' \mathrm{x_{it}} + \sigma_{\mathrm{uit}} |\mathrm{U_{it}}|}{\sigma_{\mathrm{v}}} \mathrm{p}\,(|\mathrm{U_{it}}|)\, \mathrm{d}\, |\mathrm{U_{it}}| \right) \quad (5)$$

where $\mathrm{p}\,(|U_{it}|) = 2\phi(|U_{it}|)$, and $|U_{it}| \in [0, \infty)$.[4]

While W. Greene (2010) considered simulation (with, for instance, Halton draws for U_{it}), the integral in (5) can also be approximated using the quadrature method (i.e., Gauss-Kronrod routine) which is the approach that we follow here as it is more efficient.

In the presence of sample selection, the canonical form of the Heckman model is:

$$\mathrm{y_{it}} = \beta' \mathrm{x_{it}} + \mathrm{v_{it}}$$

$$\mathrm{d_{it}^*} = \gamma' \mathrm{z_{it}} + \mathrm{w_{it}} \quad (6)$$

where d_{it}^* is a latent (unobserved) variable; z_{it} is a vector of explanatory variables; γ are parameters to be estimated; and w_{it} is an error term.

The second equation in (6) is the selection equation which can be estimated using a binary probit selection criterion $[d_{it} = 1(d_{it}^* > 0) \wedge d_{it} = 0(d_{it}^* \le 0)]$. Moreover, values of y and x are only observed when $d_{it} = 1$, and (v, w) ~ $N_2[{}^0_0, ({}^{\sigma_v^2}_{\rho\sigma_v} \; {}^{\rho\sigma_v}_{\sigma_w^2 = 1})]$; where N_2 indicates the bivariate normal distribution.

Selection models such as the one in (6) can be estimated using two methods. One method is the Heckman (1979) two-step, limited information procedure, which consists of estimating the aforementioned probit model (second equation in (6)) in a first step. In a second step, the Inverse Mills Ratio (IMR) is generated for each observation and then used as an additional explanatory variable in the linear regression estimation of the top equation in (6). However, as underlined in W. Greene (2010, p. 17), the two-step approach is inappropriate in the case of non-linear models such as the SF, and only a few studies have used the two-step estimator in the case of SF (Sipiläinen & Oude Lansink, 2005; Solís et al., 2007). Another method used

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists
IAAE
WILEY

to estimate the model in (6) is the full information maximum likelihood—FIMLE (see Maddala, 1983), which is more appropriate in the presence of inefficiency.

The combination of (1) and (6) yields

$$y_{it} = \beta' x_{it} - \sigma_{uit} |U_{it}| + v_{it}$$

$$d_{it}^{*} = \gamma' z_{it} + w_{it}$$

$$(v,\ w) \sim N_2 \left[\begin{matrix} 0 \\ 0' \end{matrix} \begin{pmatrix} \sigma_v^2 \rho \sigma_v & \\ \rho \sigma_v\ \sigma_w^2 = & 1 \end{pmatrix} \right] \tag{7}$$

where the correlation between w_{it} and v_{it}, ρ, captures the sample selection.

The full information likelihood is built up from (i) Prob(selection) × density|selection for selected observations; and (ii) Prob(non − selection) for non-selected observations. Thus, the conditional (on u_{it}) density can be written as:

$$L_{it} = d_{it} \left\{ f\left(y_{it} x_{it},\ |U_{it}|,\ w_{it} > -\gamma' z_{it}\right) P\left(w_{it} > -\gamma' z_{it}\right) \right\} + (1 - d_{it}) P\left(w_{it} \leq -\gamma' z_{it}\right) \tag{8}$$

where expression $f(y_{it}x_{it}, |U_{it}|, w_{it} > -\gamma' z_{it})P(w_{it} > -\gamma' z_{it})$ is equivalent to $\int_{-\gamma' z_{it}}^{\infty} f(v_{it}, w_{it}) dw_{it}$.[5]

From the bivariate normal distribution, we know that $w_{it}|v_{it} \sim N(\frac{\rho}{\sigma_v} v_{it},\ 1 - \rho^2)$. Then

$$\int_{-\gamma' z_{it}}^{\infty} f(v_{it}, w_{it})\, dw_{it} = f(v_{it}) \int_{-\gamma' z_{it}}^{\infty} f(w_{it}|v_{it})\, dw_{it}$$

$$= \phi(v_{it})\, \Phi \left(\frac{\frac{\rho}{\sigma_v} v_{it} + \gamma' z_{it}}{\sqrt{1 - \rho^2}} \right)$$

where Φ is the cumulative distribution of the standard normal distribution. Thus, we have

$$f\left(y_{it} x_{it},\ |U_{it}|,\ w_{it} > -\gamma' z_{it}\right) = \phi\left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) \Phi \left(\frac{\frac{\rho}{\sigma_v} \left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) + \gamma' z_{it}}{\sqrt{1 - \rho^2}} \right) \tag{9}$$

As previously, the log-likelihood function of the model specified in (7) is obtained by integrating the density in (8) over the range of $|U_{it}|$. Thus:

$$\log f\left(y_{it} x_{it}\right) = \log \int_{|U_{it}|} f\left(y_{it} x_{it},\ |U_{it}|,\ d_{it}\right) p\left(|U_{it}|\right) d\,|U_{it}| \tag{10}$$

with $|U_{it}| \in [0,\ \infty)$. Specifically, we have:

$$\log f\left(y_{it} x_{it}\right) = \log \left[d_{it} \int_{|U_{it}|} \phi\left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) \Phi \times \left(\frac{\frac{\rho}{\sigma_v} \left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) + a_{it}}{\sqrt{1 - \rho^2}} \right) d\,|U_{it}| + (1 - d_{it})\, \Phi\left(a_{it}\right) \right] \tag{11}$$

where $a_{it} = \hat{\gamma}' z_{it}$ is obtained from the first-step probit model which is similar to the two-step Heckman model, and $|U_{it}| \in [0,\ \infty)$.

Under this simplification, maximizing the log-likelihood function specified in (11) is a hybrid two-step Limited Information Maximum Likelihood (LIML) estimation. As underlined in W. Greene (2010), the non-selected observations ($d_{it} = 0$) do not contribute information about the parameters, and hence the log-likelihood function to be maximized is simplified to

$$\log f\left(y_{it} x_{it}\right) = \log \left[\int_{|U_{it}|} \phi\left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) \Phi \left(\frac{\frac{\rho}{\sigma_v} \left(y_{it} - \beta' x_{it} + \sigma_{uit} |U_{it}|\right) + a_{it}}{\sqrt{1 - \rho^2}} \right) d\,|U_{it}| \right] \tag{12}$$

with $|U_{it}| \in [0,\ \infty)$.

In contrast to Bravo-Ureta et al. (2020) where the selection mechanism is considered at a single base period (a_i is constant over time), in our case the selection mechanism is not fixed over time. Therefore, our model is estimated as a pooled model and Murphy and Topel (2002)'s correction is applied to obtain the appropriate standard errors (see Lai, 2015, p. 109, formula [14] for more details).

2.1.2 | Efficiency of observations

The estimation of each observation's efficiency follows the line of Jondrow et al. (1982) but is adapted to account for sample selection. Following W. Greene (2010), we have:

$$p(u_{it} \epsilon_{it}) = \frac{p\left(u_{it},\ \epsilon_{it}\right)}{p(\epsilon_{it})} = \frac{p\left(\epsilon_{it} u_{it}\right) p\left(u_{it}\right)}{\int_{u_{it}} p\left(\epsilon_{it} u_{it}\right) p\left(u_{it}\right) du_{it}} \tag{13}$$

where $\epsilon_{it} = v_{it} - u_{it}$.

Therefore

$$E[u_{it}|\epsilon_{it}] = \frac{\int_{u_{it}} u_{it} p\left(\epsilon_{it} u_{it}\right) p\left(u_{it}\right) du_{it}}{\int_{u_{it}} p\left(\epsilon_{it} u_{it}\right) p\left(u_{it}\right) du_{it}} \tag{14}$$

with p() the probability and $u_{it}=\sigma_{uit}\,|U_{it}|$. Therefore $u_{it}\in[0,\ \infty)$.

Thus, the denominator of formula (14) is obtained as the predicted value of $f(y_{it}x_{it})$ in Equation (12). In our case, the numerator's integral is solved with the parameters obtained from maximizing (12). Finally, the efficiency is computed as $\exp[-\hat{E}[u_{it}|\epsilon_{it}]]$.

2.2 | Sample selection correction in the LCSFM case

2.2.1 | Likelihood construction

As underlined in Orea and Kumbhakar (2004), in the presence of technological differences in the SF framework discussed earlier, the results obtained may be biased since the unobserved technological heterogeneity might be confounded with producer-specific inefficiency. To handle this technological heterogeneity, one can assume a discrete (finite) mixture of several technologies that approximates to the continuous random parameters model (W. Greene, 2005).

Let's assume that the data can be segmented into J latent classes. The LCSFM is a single-stage approach where the probability of class membership and the mixture of technologies are simultaneously estimated. Each observation i in time t belongs to class j with a prior probability (which can also be viewed as the mixture weight) parameterized using a multinomial logit function, as is commonly done in latent class analysis (W. H. Greene, 2001):

$$\Pi_{itj}\ \left(\lambda|q_{ih(t)}\right)=\frac{\exp(\lambda'_j q_{ih(t)})}{\sum_{m=1}^{J}\exp(\lambda'_m q_{ih(t)})}\quad j=1,\dots,J;\ \lambda_J=\ 0;\ h=1,2,3;\ t=1,\ \dots,T$$

$$0\le\Pi_{itj}\ \left(\boldsymbol{\lambda}|q_{ih(t)}\right)\le 1;\ \sum_j\Pi_{itj}\ \left(\boldsymbol{\lambda}|q_{ih(t)}\right)=\ 1 \tag{15}$$

where $\mathbf{q}_{ih(t)}$ is the vector of the period- and firm-specific variables that explains the probability of belonging to one class or another (separating variables); $\boldsymbol{\lambda}_j$ is a vector of unknown coefficients for the latent class j (relative to the base class J), and $\boldsymbol{\lambda}\ =(\boldsymbol{\lambda}_1,\ \boldsymbol{\lambda}_2,\ \dots,\ \boldsymbol{\lambda}_J)$ Three periods h are considered, corresponding to the implementation in France of major reforms of the CAP: the first period is 2002–2006, the second 2007–2013, and the third 2014–2016. Practically, for each period h and each farm i, the variable q is averaged over the time period, which makes the probability of belonging to a class constant over the period h.

Equation (15) shows one empirical contribution of our modeling: the prior probabilities are period-dependent instead of being fixed through time, as is usually the case in the literature. For example, in W. H. Greene (2002), Orea and Kumbhakar (2004), and W. Greene (2005), it is assumed a priori that a firm permanently belongs to a specific class, and the prior probabilities represent the analyst's knowledge uncertainty and not the state of nature. With this in mind, some authors (Alvarez & del Corral, 2010; Orea & Kumbhakar, 2004) recommend using firm averages to construct time-independent separating variables. However, we consider that specific periods associated with important changes in the CAP may be significant factors. Therefore, as the model estimated is a pooled LCSFM, for each observation i and each time t, the separating variables are equal to the average (of the observation) over the specific period h (repeatedly). In other words, here farms are a priori allowed to move from one class to another class only twice, depending on the periods.

Under the LCSFM framework and endogenous sample selection, the production technology can be described as follows:

$$y_{it}=\ \beta_j'\ x_{it}-\sigma_{uit,j}\left|U_{it,j}\right|+v_{it,j}\ j=1,\dots,J$$

$$d^*_{it}=\ \gamma'\ z_{it}+w_{it}$$

$$\left(v_j,\ w\right)\sim N_2\left[\begin{matrix}0\\0'\end{matrix}\left(\begin{matrix}\sigma^2_{v,j} & \rho\sigma_{v,j}\\ \rho\sigma_{v,j} & \sigma^2_w=\ 1\end{matrix}\right)\right]\ j=1,\dots,J \tag{16}$$

Since we consider heterogeneity in the production technologies only, the sample selection equation (second equation in (16)) is the same as the one for the case without production heterogeneity (see Equation (7)).

The (conditional) log-likelihood function associated with the formulation in (16) is

$$LF_{it|j}\ \left(\theta_j|x_{it},H_{it},\ a_{it}\right)=\int_{|U_{it,\,j}|}\phi\left(y_{it}-\beta_j'x_{it}+\sigma_{uit,j}\left|U_{it,j}\right|\right)\Phi\left(\frac{\frac{\rho_j}{\sigma_{v,\,j}}\left(y_{it}-\beta_j'x_{it}+\sigma_{uit,\,j}\left|U_{it,\,j}\right|\right)+a_{it}}{\sqrt{1-\rho_j^2}}\right)d\left|U_{it,\,j}\right| \tag{17}$$

where $\theta_j=\ (\beta_j,\ \delta_j,\ c_j,\ \rho_j)$ are the parameters to be estimated for each class j, and $|U_{it,j}|\in[0,\ \infty)$.

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists

The unconditional likelihood of observing i in time t is then obtained by averaging over classes:

$$LF_{it}\left(\Theta|x_{it},H_{it},a_{it},q_{ih(t)}\right)=\sum_{j=1}^{J}\Pi_{itj}\left(\lambda|q_{ih(t)}\right)\times LF_{it|j}\left(\theta_j|x_{it},H_{it},a_{it}\right) \tag{18}$$

where $\mathbf{\Theta} = (\theta_1, \theta_2, \ldots, \theta_J, \lambda)$. The log-likelihood of the LCSFM (log L) in our case can be written as:

$$\log L=\log\left(\prod_{i=1}^{N}\prod_{t=1}^{T}LF_{it}\left(\mathbf{\Theta}|x_{it},H_{it},a_{it},q_{ih(t)}\right)\right)$$

$$\log L=\sum_{i=1}^{N}\sum_{t=1}^{T}\log\left(\sum_{j=1}^{J}\Pi_{itj}\left(\lambda|q_{ih(t)}\right)\times LF_{it|j}\left(\theta_j|x_{it},H_{it},a_{it}\right)\right) \tag{19}$$

This log-likelihood is maximized using the conventional BFGS (Broyden-Fletcher-Goldfarb-Shanno) method. The standard errors are obtained using Murphy and Topel (2002)'s correction.

Using the Bayes conditional probability theorem, we can compute the posterior probability of belonging to class j in time t as:

$$P\left(class = j|x_{it},H_{it},a_{it},q_{ih(t)}\right) = \frac{\Pi_{itj}\left(\lambda|q_{ih(t)}\right)\times LF_{it|j}\left(\theta_j|x_{it},H_{it},a_{it}\right)}{\sum_{m=1}^{J}\Pi_{itm}\left(\lambda|q_{ih(t)}\right)\times LF_{it|m}\left(\theta_m|x_{it},H_{it},a_{it}\right)} \tag{20}$$

Each observation can be assigned to a specific class considering the largest posterior probability. As underlined in Parmeter and Kumbhakar (2014), some observations may have a probability of belonging to a specific class close to unity and, therefore, it is consistent with using the technological parameters of this class for these observations. However, other observations may have non-unity probabilities of belonging to different classes. Nevertheless, in our case, the parameters considered for each observation are the ones associated with the class that has the highest (posterior) probability. The number of classes to consider can be based on a comparison of the Akaike Information Criterion (AIC) across models, as suggested by Orea and Kumbhakar (2004).[6]

2.2.2 | Efficiency of observations

For each class frontier, the inefficiency is estimated using Equations (13) and (14) for each technology. As pointed out in W. Greene (2010), the inefficiency scores obtained in this way are very close to the ones obtained using the Jondrow et al. (1982) formula.

Finally, as mentioned above, an additional contribution of our article is to extend the LCSFM to the incorporation of heteroscedasticity in the one-sided error term. In other words, inefficiency effects are estimated in the same single stage as the other LCSFM parameters.

3 | DATA AND EMPIRICAL MODEL

Our study uses an unbalanced panel dataset of French dairy producers obtained from the annual farm-level French Farm Accountancy Data Network (FADN) during 2002–2016. The FADN database, managed by the French Ministry of Agriculture, includes accountancy data of representative commercial farms in France. The sample of dairy farms used here consists of 15,623 observations in total in 2002–2016.[7]

Following the literature (Bradfield et al., 2021; Kellermann & Salhofer, 2014; Skevas et al., 2018), in the production technology specification, two outputs are used; namely, the quantity of milk produced (Y_1; in tons) and the other output value (Y_2; in constant Euros)[8], and five inputs (X_1, X_2, X_3, X_4, X_5) which are: utilized agricultural area (UAA) (in hectares - ha); total labor (in full time equivalent annual working units); herd size (in livestock units)[9]; intermediate inputs (in constant Euros)[10]; and fixed assets excluding land and herd (in constant Euros). In addition, to control for exogenous factors in the production process, we include year fixed effects (*D*) and a dummy indicating whether the farm is located in a less favored area (LFA), assuming that being located in LFA indicates low soil productivity.[11]

We specify the technology as a negative Translog output distance function, exploit its homogeneity in output quantities, and assume $\ln D\ (Y_{it}, X_{it}, t, LFA_i) = -u_{it}$,[12] similarly to Dong et al. (2016), Orea et al. (2015), Pérez-Méndez et al. (2020) as follows:

$$\begin{aligned}\ln Y_{1it} &= \beta_0+\sum_{k=1}^{5}\beta_k \ln X_{kit}+\eta \ln\frac{Y_{2it}}{Y_{1it}}\\&+\frac{1}{2}\sum_{k=1}^{5}\sum_{l=1}^{k}\beta_{kl}\ln X_{kit}\ln X_{lit}+\frac{\kappa}{2}\left(\ln\frac{Y_{2it}}{Y_{1it}}\right)^2\\&+\sum_{k=1}^{5}\tau_k \ln X_{kit}\ln\frac{Y_{2it}}{Y_{1it}}+\sum_{t=2003}^{2016}\\&\beta_t D_t+\beta_{LFA}LFA_i+v_{it}-u_{it}\end{aligned} \tag{21}$$

with $\beta_{kl} = \beta_{lk}$ for $k, l = 1, \ldots, 5$.

Two separating variables (q_1, q_2) are used in the LCSFM to identify the farm classes: (i) the stocking rate, calculated as the number of livestock units per ha of UAA. This variable is an indicator of grazing pressure: the higher this variable, the more intensive is the technology; and (ii) the share of permanent grassland in the UAA, which may shed light on the available pastures for animal feed: the higher the share, the more extensive is the production technology. These two separating variables are not included in the production function since they proxy practices and are not inputs nor exogenous controls. In the existing literature, similar separating variables are used (see, for instance, Alvarez & del Corral, 2010; Kellermann & Salhofer, 2014).

As explained above, one main modeling contribution is that we simultaneously investigate the drivers of inefficiency while correcting for selectivity in the LCSFM. Five variables ($H_1, \ldots, H_5$) are used to model the one-sided error variance. Based on the large literature on farms' technical efficiency (Bonfigl et al., 2020; Cabrera et al., 2010; Dakpo, Latruffe, et al., 2021; del Corral et al., 2011; Dong et al., 2016; Latruffe et al., 2017; Minviel & Latruffe, 2017), we include:

- a farmer's socio-demographic characteristics: farmer's age (H_1); and a low education dummy (H_2) taking the value one if the farmer has a low level of education (that is, either no education or primary education) and zero if the farmer has a high level of education (secondary education or above);
- the resort to external inputs: the share of hired labor in total labor (H_3); and the share of rented area in UAA (H_4);
- public support reliance: operational subsidies per ha of UAA (H_5); these are provided to farms under the framework of the CAP and include fully decoupled subsidies in the form of the Single Farm Payment, subsidies coupled to the acreage of specific crops and to the headage of specific livestock, subsidies received from adopting AESs, and subsidies received for being located in a LFA.

Regarding the first-step probit selection equation, as mentioned in the introduction, a large body of the literature has examined AESs adoption determinants (Coyne et al., 2021; Defrancesco et al., 2008; Lefebvre et al., 2020). Based on this literature and also considering that AESs may induce changes in farmers' practices (Arata & Sckokai, 2016), thus preventing the use of several farm characteristics, the following simplified model is used:

$$d_{it}^{*}=\gamma_0+\sum_{m=1}^{3}\gamma_m z_{mit}+\sum_{r=1}^{20}\zeta_r R_{rit}+w_{it} \tag{22}$$

where z_1 is the milk price (in Euros per ton of milk),[13] z_2 and z_3 are farmers' characteristics namely age and low education dummy respectively ($z_2=H_1$; $z_3=H_2$), and R represents regional dummies (21 regions in total).

Table 1 displays some of the descriptive statistics of the variables. Overall, 34% of our sample observations have adopted AES, and we will call them the sub-sample of AESs adopters. On average, output levels, milk yield, herd size, and intermediate consumption are higher for farmers who have not adopted AESs, who we call the sub-sample of AESs non-adopters. By contrast, UAA and milk price are higher for the sub-sample of AESs adopters on average. About 75% of the AESs adopters are located in LFAs, while this proportion is 36% for non-adopters. In terms of the separating variables, the stocking rate is higher for AESs non-adopters' sub-sample. At the same time, the share of permanent grassland in UAA is lower, indicating more intensive practices for this sub-sample than for the sub-sample of AESs adopters. In terms of the drivers of inefficiency, the descriptive statistics are relatively similar for both sub-samples. Finally, within each sub-sample, a high heterogeneity is observed, with high values for the coefficients of variation (>25% for most variables), suggesting the relevance of using a latent class model for each sub-sample.

4 | RESULTS

The main results of the estimation of the first-step probit model in (22) are displayed in Table 2 (full results are presented in Table A1 of the Supplementary online appendix). Results show that the probability of adopting AESs increases with milk price. Younger farmers and low-educated farmers have a higher probability of adoption than older farmers and more highly educated farmers. Although the result regarding education is not in line with the literature (Siebert et al., 2006), our first-step probit's explanatory variables are highly significant. The rate of good predictions is 72.12%, suggesting that our selection model is valid.

We now turn to the main results; namely those for the LCSFM that has been estimated separately for each sub-sample (AESs adopters; AESs non-adopters), accounting for the sample selection parameter ρ. As usual for the Translog functional form, explanatory input and output variables ($x_1, x_2, x_3, x_4, x_5, y_1, y_2$) are scaled by their (geometric) means. Therefore, the first-order coefficients represent distance elasticities at the sample (geometric) mean. In each sub-sample, farms are categorized into two classes.[14] Moreover, in addition to the LCSFM with two classes, for comparison purposes, we have estimated a standard sample-selection stochastic frontier model (i.e.,

TABLE 1 Descriptive statistics of the data during 2002–2016

Variables	Full sample Mean	Standard deviation	Median	1th percentile	99th percentile	Sub-sample of adopters of AESs Mean	Standard deviation	Median	1th percentile	99th percentile	Sub-sample of non-adopters of AESs Mean	Standard deviation	Median	1th percentile	99th percentile
Production function															
Milk production (tons): Y_1	338.7	202.0	292.1	63.9	1,032.7	289.6	171.8	246.8	62.6	914.21	364.1	211.6	319.9	65.2	1069.8
Other output value (thousand Euros): Y_2	64.1	52.9	48.9	5.1	249.8	55.2	46.4	41.1	5.8	225.2	68.8	55.5	53.8	4.7	259.6
UAA (hectares): X_1	89.8	50.5	77.0	21.1	259.8	94.8	51.5	81.2	24.7	264.2	87.2	49.8	74.9	20.4	251.0
Total labor (annual working units): X_2	1.9	.9	2.0	1.0	4.6	1.9	.9	2.0	1.00	4.84	1.9	.9	2.0	1.00	4.52
Herd size (livestock units): X_3	99.0	55.1	86.6	24.2	283.3	93.0	51.8	80.2	24.7	259.8	102.1	56.5	90.2	23.8	294.5
Intermediate consumption (thousand Euros): X_4	96.3	62.0	81.7	17.2	305.1	81.0	52.7	66.6	15.9	261.9	104.2	64.9	89.3	18.6	317.3
Fixed assets (thousand Euros): X_5	186.5	159.3	146.7	3.9	748.7	191.6	153.8	152.0	5.5	692.7	183.8	162.0	143.9	3.2	764.5
Farm in LFA (dummy: 1 if yes, 0 if not)	.49	.50	.00	.00	1.00	.75	.43	1.00	.00	1.00	.36	.48	.00	.00	1.00

(Continues)

TABLE 1 (Continued)

Variables	Full sample Mean	Standard deviation	Median	1th percentile	99th percentile	Sub-sample of adopters of AESs Mean	Standard deviation	Median	1th per-centile	99th per-centile	Sub-sample of non-adopters of AESs Mean	Standard deviation	Median	1th percentile	99th percentile
Separating variables															
Stocking rate (livestock units per ha of UAA): $q1$	1.2	.4	1.1	.5	2.5	1.0	.4	1.0	.40	2.15	1.2	.4	1.2	.56	2.57
Share of permanent grassland in UAA: $q2$	.40	.32	.37	.00	1.00	.57	.34	.63	.00	1.00	.31	.28	.28	.00	1.00
First-step probit model															
Milk price (Euros per ton): z_1	339.9	51.4	329.6	265.5	516.2	348.2	60.6	334.8	263.1	530.1	335.5	45.3	327.8	266.8	510.6
Inefficiency drivers															
Farmer's age (years): $H_1 = z_2$	47.1	8.5	48.0	27	63.0	46.4	8.4	47.0	28	63	47.4	8.6	48.0	26	63
Farmer's low education (dummy: 1 if yes, 0 if not): $H_2 = z_3$	.24	.43	.00	.00	1.00	.25	.43	.00	.00	1.00	.24	.43	.00	.00	1.00
Share of hired labor in total labor: H_3	.06	.14	.00	.00	.59	.06	.13	.00	.00	.52	.07	.15	.00	.00	.61
Share of rented area in UAA: H_4	.79	.29	.96	.00	1.00	.79	.28	.94	.00	1.00	.80	.29	.97	.00	1.00
Operational subsidies per ha of UAA (Euros): H_5	378.5	117.2	365.8	149.8	726.6	381.1	116.6	371.1	161.9	718.3	377.1	117.4	363.0	144.2	727.8
Number of observations	15,436					5,274					10,162				

TABLE 2 First-step probit model (determinants of AESs adoption) on the full sample

Variables	Estimates and significance
Milk price (z_1)	.0009***
Farmer's age (z_2)	−.015***
Farmer's low education dummy (z_3)	.091***
Rate of good predictions	72.12%
Pseudo R^2	.186

Notes: *, **, *** indicate significance at the 10%, 5%, 1% level, respectively. Regional dummies are included in the regression, and the results are shown in the appendix.

with one single class) for each of the two sub-samples of farms.

The main results of the estimation for the single-class model (a) and the LCSFM with two classes (b) are presented in Table 3 for the AESs adopters' sub-sample. Similarly, Table 4 presents the results for the AESs non-adopters sub-sample (full results are available in Tables A2 and A3, respectively, in the Supplementary appendix).

First, Tables 3 and 4 show that the selectivity parameter (ρ) is significant, confirming the presence of sample selection bias and the necessity to correct for the bias with the first-step probit model. Second, for each sub-sample, results show differences in coefficients and elasticities between the single-class and the two-class models, highlighting the importance of accounting for technological heterogeneity in the frontier estimation.

In order to assess the reliability of the results in particular of the efficiency estimates (see Henningsen & Henning, 2009), we report in Table A6 the share of observations that satisfy the monotonicity conditions for models with sample selection. Monotonicity conditions are more frequently violated within the sub-sample of AES adopters than in the sub-sample of non-adopters. Within each sub-sample, those conditions are more fulfilled in the intensive class than in the extensive class. Overall, depending on the sub-sample and the class, the share of observations that satisfy all the monotonicity conditions ranges between 39% and about 87%.

Focusing firstly on the estimation for AES adopters' sub-sample (Table 3), the results regarding the separating variables indicate that the probability of belonging to class 1 is negatively associated with a higher stocking rate and positively associated with a higher share of permanent grassland in UAA. Based on these results, class 1 can be deemed extensive, while class 2 is intensive. The estimation results of the distance function indicate that for both classes, the highest distance elasticity is associated with the intermediate inputs, followed by herd size. While LFA is negatively

TABLE 3 Estimated coefficients of the latent Translog production frontier for the AESs adopters' sub-sample: Comparison of (a) the single class model, and (b) the latent class model with two classes—Estimation with sample selection

		Latent class model with two classes (b)	
Variables	Standard model (single class) (a)	Class 1 (extensive)	Class 2 (intensive)
Production function: Elasticities at sample geometric mean			
log(X_1: UAA)	.025**	.186***	.029**
log(X_2: total labor)	.122***	.071***	.142***
log(X_3: herd size)	.231***	.204***	.24***
log(X_4: intermediate consumption)	.554***	.485***	.5***
log(X_5: fixed assets)	.06***	.066***	.052***
log(Y_2/Y_1: other output/milk production)	−.25***	−.268***	−.239***
Farm in LFA dummy	−.079***	−.016	−.069***
Sample selection parameter			
Rho (ρ)	.326***	−.424***	.541***
Separating variables			
Stocking rate: q_1	–	−1.07***	–
Share of permanent grassland in UAA: q_2	–	4.286***	–
Inefficiency drivers			
Farmer's age: $H_1/100$	−.344	−1.793***	6.61*
Farmer's low education dummy: H_2	.274***	.431***	−.431
Share of hired labor in total labor: H_3	−.446**	−.579*	−15.069
Share of rented area in UAA: H_4	−.155	.023	−3.854***
Operational subsidies per ha of UAA: $H_5/1000$	1.571***	1.663***	3.709**
Log-likelihood	−2,601.22	−2,197.57	
Scale elasticity at sample means	.99	1.01	.96
Average efficiency	.86	.85	.98
Average posterior probability	1	.79	.83
Number of observations	5,274	2,759	2,515

Note: *, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

associated with productivity for the intensive class, it is non-significant for the extensive class. Comparing the distance elasticities (at sample means) between both classes reveals that land elasticity in the extensive class is about sixfold that in the intensive class. On the other hand, labor

TABLE 4 Estimated coefficients of the latent Translog production frontier for the AESs non-adopters' sub-sample: Comparison of (a) the single class model, and (b) the latent class model with two classes—Estimation with sample selection

Variables	Standard model (single class) (a)	Latent class model with two classes (b): Class 1 (extensive)	Class 2 (intensive)
Production function: Elasticities at sample geometric mean			
log(X_1: UAA)	.033***	−.015	.063***
logXx_2: total labor)	.146***	.124***	.147***
log(X_3: herd size)	.227***	.29***	.255***
log(X_4: intermediate consumption)	.567***	.605***	.461***
log(X_5: fixed assets)	.058***	.066***	.057***
log(Y_2/Y_1: other output/milk production)	−.273***	−.303***	−.241***
Farm in LFA dummy	−.053***	−.034***	−.043***
Sample selection parameter			
Rho (ρ)	.083**	−.256***	.118*
Separating variables			
Stocking rate: q_1	–	.005	–
Share of permanent grassland in UAA: q_2	–	4.186***	–
Inefficiency drivers			
Farmer's age: $H_1/100$	.836***	.093	.29
Farmer's low education dummy: H_2	.41***	.271***	.87***
Share of hired labor in total labor: H_3	−.161	−.175	.118
Share of rented area in UAA: H_4	−.115*	−.042	−.643**
Operational subsidies per ha of UAA: $H_5/1000$	−.093	.562***	−3.217***
Log-likelihood	−11,872.22	−11,037.16	
Scale elasticity at sample means	1.03	1.07	.98
Average efficiency	.87	.84	.95
Average posterior probability	1	.79	.79
Number of observations	10,162	4,495	5,667

Note: *, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

elasticity is twice as high in the intensive class. For the other inputs, the results are quite close.

The extensive class exhibits slightly increasing returns to scale (1%) at the sample means, while for the intensive class, the returns to scale are decreasing (-4%). The average efficiency score is higher for the intensive class. The efficiency score for the intensive class is 98% on average (almost zero inefficiencies in the class), while it is 85% in the extensive class.

Further comparisons of the two classes are presented in Table 5. The class which is the more efficient and more intensive in terms of stocking rate and share of permanent grassland in UAA (class 2), is further characterized by more intensive practices than the other class: higher consumption of intermediate inputs (fertilizers, pesticides, energy, concentrate feed, veterinary expenses), higher share of fodder maize in UAA, and lower amount of AES subsidies per hectare of UAA, compared to the extensive class. Farms in the intensive class are, on average, larger in terms of output produced and fixed assets than farms in the extensive class, and receive more operating subsidies per hectare of UAA but a lower milk price.

Going back to Table 3, we now comment on the inefficiency determinants for each class of the sub-sample of AESs adopters. Negative signs of the coefficients indicate a negative association with inefficiency; that is, a positive impact on efficiency. In contrast, positive signs indicate a positive association with inefficiency and hence a negative association with efficiency. In both classes, the level of operational subsidies per hectare of UAA is negatively associated with efficiency, as it is also in the standard (single class) model. Specific to each class are the associations between efficiency and of farmer's age, low education dummy, hired labor's share, and rented area's share. While farmer's age is positively associated with farm's technical efficiency in the extensive class, the relationship is reversed in the intensive class. The low education variable is negatively associated with the extensive class's technical efficiency and is non-significant in the intensive class. For this latter class, the share of rented area is positively associated with the level technical efficiency while in the extensive class, it is the share of hired labor that is positively associated with technical efficiency; the other relationships being non-significant.

It is important to note here an illustration of the need to account for technological heterogeneity to generate targeted policy recommendations. For example, in the single-class model, low education is negatively associated with the efficiency; however, when specific class models are estimated, this negative relationship holds for farms in the extensive class only. The only consistent result across the single class and both classes, is the negative association between operational subsidies and technical efficiency.

AGRICULTURAL ECONOMICS
The Journal of the International Association of Agricultural Economists
IAAE
WILEY

TABLE 5 Average farm characteristics for each class for the AES adopters' sub-sample

Variables	Class 1 (extensive)	Class 2 (intensive)
Main production (tons of milk): Y_1	253.00	329.84
Other output (thousand Euros): Y_2	46.07	65.27
UAA (ha): X_1	97.80	91.57
Total labor (annual working units): X_2	1.78	1.93
Herd size (livestock units): X_3	91.55	94.62
Intermediate consumption (thousand Euros): X_4	72.43	90.33
Fixed assets (thousand Euros): X_5	186.85	196.80
Farm in LFA (dummy: 1 if yes, 0 if not): LFA	.88	.60
Stocking rate (livestock units per ha of UAA): q_1	.95	1.11
Share of permanent grassland in UAA: q_2	.81	.31
AES subsidies per hectare of UAA (Euros): q_4	70.48	60.84
Farmer's age (years): H_1	46.80	46.00
Farmer's low education (dummy: 1 if yes, 0 if not): H_2	.22	.27
Share of hired labor in total labor: H_3	.05	.07
Share of rented area in total UAA: H_4	.79	.78
Operational subsidies per ha of UAA (Euros per ha): H_5	358.18	406.30
Concentrated feed expenses per livestock unit (Euros)	250.32	257.69
Veterinary expenses per livestock unit (Euros)	42.49	51.05
Fertilizers expenses per ha of UAA (Euros)	58.22	94.01
Pesticides expenses per ha of UAA (Euros)	12.91	36.60
Energy expenses per ha of UAA (Euros)	34.37	42.44
Share of fodder maize in UAA (%)	.04	.13
Milk price (Euros per ton)	358.36	337.06
Milk volume per milking cow (tons per cow)	5.31	6.16

TABLE 6 Average farm characteristics for each class for the AES non-adopters' sub-sample

Variables	Class 1 (extensive)	Class 2 (intensive)
Main production (tons of milk): Y_1	313.88	403.94
Other output (thousand Euros): Y_2	60.85	75.04
UAA (ha): X_1	89.87	85.16
Total labor (annual working units): X_2	1.83	1.92
Herd size (livestock units): X_3	105.18	99.72
Intermediate consumption (thousand Euros): X_4	95.84	110.83
Fixed assets (thousand Euros): X_5	185.76	182.26
Farm in LFA (dummy: 1 if yes, 0 if not): LFA	.42	.31
Stocking rate (livestock units per ha of UAA): q_1	1.23	1.25
Share of permanent grassland in UAA: q_2	.52	.15
Farmer's age (years): H_1	48.06	46.92
Farmer's low education (dummy: 1 if yes, 0 if not): H_2	.26	.22
Share of hired labor in total labor: H_3	.07	.07
Share of rented area in total UAA: H_4	.81	.79
Operational subsidies per ha of UAA (Euros per ha): H_5	363.77	387.75
Concentrated feed expenses per livestock unit (Euros)	258.72	299.50
Veterinary expenses per livestock unit (Euros)	44.69	57.32
Fertilizers expenses per ha of UAA (Euros)	112.21	133.33
Pesticides expenses per ha of UAA (Euros)	45.86	66.39
Energy expenses per ha of UAA (Euros)	41.61	49.77
Share of fodder maize in UAA (%)	.16	.24
Milk price (Euros per ton)	343.28	329.36
Milk volume per milking cow (tons per cow)	5.75	6.95

Looking at the sub-sample of farmers who have not adopted AESs (Table 4), the probability of belonging to class 1 is positively related to a higher share of permanent grassland in UAA and non-significantly related to the stocking rate. These results suggest that class 1 is extensive, while class 2 is intensive. This is confirmed by further examination of the characteristics of each class, as shown in Table 6. Table 6 shows a very close value in terms of stocking rate for the two classes, but a lower share of permanent grassland in UAA and higher consumption of intermediate inputs (fertilizers, pesticides, energy, concentrate feed, veterinary expenses) for the intensive class 2 compared to the extensive class 1. A conclusion from Table 6 is that using the stocking rate and the share of permanent grassland in UAA as separating variables in the LCSFM is sufficient to generate two contrasting classes in terms of intensive or extensive practices.

The estimation results of the distance function in Table 4 show that, similarly to the sub-sample of AESs adopters, the input with the highest elasticity are intermediate inputs, followed by herd size. A divergence is that the elasticity of land is non-significant for the extensive class in the AESs non-adopters' sub-sample (Table 4).This class exhibits moderate increasing returns to scale (1.07, see Table 4) while the intensive class has slightly decreasing returns to scale (.98). Similarly, to the sub-sample of AESs adopters, in the sub-sample of AESs non-adopters, the intensive class is, on average, more efficient (with a 95% efficiency average compared with 84% for the extensive class). A further look at the class characteristics (Table 6) reveals that farms in the intensive class on average receive a higher amount of operational subsidies per ha of UAA but a lower milk price, similar to the case of AESs adopters (Table 5).

Finally, going back to the estimation results in Table 4, and similarly to the case of the extensive class in AESs adopters in Table 3, for AESs non-adopters low education is negatively associated with efficiency for both classes. In contrast, operational subsidies per ha of UAA are negatively associated with efficiency for the extensive class only, and positively associated with efficiency for the intensive class, while it is not significant in the single class model. Here again the results underline the discrepancy in the association with of efficiency drivers depending on whether heterogeneity is taken into account, and depending on the class. The association between of the share of rented area and efficiency is positive for the intensive class, similarly to the intensive class in the case of AESs adopters. The share of hired labor has no significant relationship in either class.

To assess the importance of accounting for sample selection in the LCSFM, models without selection correction have been estimated. The corresponding likelihood and subsequent functions can be found in Parmeter and Kumbhakar (2014). Results are available in the Supplementary appendix. In the case of AESs adopters, comparing results without selection correction (Table A4 in the Supplementary appendix) and results with selection correction (Table 3), shows that similar classes are identified. The same conclusion is made when comparing results without selection correction (Table A5 in the Supplementary appendix) and results with selection correction (Table 4) for the sub-sample of AESs non-adopters. For both sub-samples, the separating variables still indicate that class 1 is the extensive one, with low average efficiency.

However, the inefficiency drivers show some contrasting impacts. Most notable is the difference regarding the association between of the share of hired labor and the efficiency of the extensive class for the sub-sample of AESs adopters: while the share of hired labor has no significant relationship without selection correction (Table A7), it is positively associated with efficiency in the model with selection correction (Table 3). Hence, not accounting for sample selection of AES adoption modifies some conclusions related to the inefficiency drivers.

5 | CONCLUSION

This article's objective was to assess farms' technical efficiency accounting for their production heterogeneity and simultaneously accounting for whether or not farms have adopted AESs. Considering the latent technologies, and to correct the potential endogeneity associated with the adoption of AESs, we have extended the LCSFM with sample selection correction, which is our major contribution.[15] We have estimated a first-step selection probit, and in a second step we have estimated the LCSFM on two separate sub-samples (the AESs adopters and the AESs non-adopters) including a term for correcting sample selection obtained in the first step. A second main contribution is that we have accounted for heteroscedasticity with the estimation of inefficiency effects in the LCSFM. The application was to FADN dairy farms in France during 2002–2016. The separating variables (namely, stocking rate and share of permanent grassland in UAA) to account for production heterogeneity in the latent class model have been constructed to allow a priori farms to change class only twice. We used the average of each separating variable for specific periods of CAP implementation in France: 2002–2006, 2007–2013, and 2014–2016.

For each of the two sub-samples (AESs adopters and AESs non-adopters) we have identified one class with intensive technology and one class with extensive technology. For both sub-samples, the farms in the intensive class rely less on permanent grassland and more on external inputs such as pesticides, fertilizers and concentrated feed, than the farms in the extensive class. However, on average, for both sub-samples the intensive class is more efficient, receives more operational subsidies per ha of UAA, but is paid a lower milk price, than the extensive class.

Our findings therefore show that there exists a heterogeneity in terms of more or less intensive technology within farms that have adopted AESs, as well as within farms that have not adopted AESs, with four classes in total, from the most extensive (the extensive class in the AESs adopters sub-sample) to the most intensive (the intensive class in the AESs non-adopters sub-sample). From a policy point of view, these results could help better target the CAP incentives provided to farmers to adopt environmentally-friendly practices: obviously, the existing AES are not suited to the sub-sample of non-adopters of

AES, and new AES could be designed in a way that they are specifically targeted to the intensive class of the sub-sample of non-adopters of AES, with our methodological approach being one possibility to identify such class.

Further insights for policy design are provided with the results regarding inefficiency. The main divergence in findings between the LCSFM for the sub-sample of AESs adopters and the LCSFM for the sub-sample of AESs non-adopters relates to the association between operational subsidies per ha and efficiency. First, in the case of AESs non-adopters, association between these subsidies and efficiency for the full sub-sample is non-significant, while it is significant for each class identified: negative association with efficiency for the extensive class, and positive association with efficiency for the intensive class.

Second, the association between subsidies and farms' efficiency in each class differs depending on whether farms have or have not adopted AESs. In both sub-samples (AESs adopters and AESs non-adopters), operational subsidies per ha have a negative association with efficiency for the extensive class. However, while the association with efficiency is also negative for the intensive class of the AESs adopters, the association with efficiency is positive for the intensive class of the AESs non-adopters.

Third, more evidence in support of this comes from the comparison of the results obtained in the LCSFM accounting for sample selection with results obtained in the LCSFM without sample selection: some results regarding inefficiency drivers, for example regarding the association with the share of hired labor in total labor, are different depending on whether sample selection is accounted for or not. Therefore, modeling production heterogeneity as precisely as possible could help better target the CAP or other policies aimed at improving farms' efficiency. For example, our results indeed show that CAP subsidies are negatively associated with the efficiency of specific groups of farms. These groups can be identified based on their technology and our methodological approach. Hence, specific subsidies could be designed specifically for each group of farms, directly with operational subsidies, or indirectly through subsidies linked to drivers identified in our models, for example, hired labor.

Our results therefore underline that, from a policy point of view, accounting for heterogeneity both in terms of intensive and extensive technology, and in terms of the adoption of AESs, could help better target the CAP. In particular, the divergence observed in terms of the association between efficiency and operational subsidies per ha supports Minviel and Latruffe (2017) who stressed that the association between subsidies on farms' technical efficiency depends on modeling.

From a methodological point of view, we have shown that it is possible to account for sample selection in LCSFM and at the same time assess inefficiency effects. Future research could investigate, reversely, the role of inefficiency in the selection equation while also accounting for production heterogeneity. Kumbhakar et al. (2009) and Latruffe and Nauges (2014) gave evidence that farms' technical efficiency influences their choice of whether to convert to organic farming or remain conventional. Kumbhakar et al. (2009) found that inefficiency decreases the probability of organic farming, while Latruffe and Nauges (2014) found that the direction of the effect depends on the farm size and main production. Separating farms based on the degree of environmental-friendliness of their farming practices with a LCSFM may provide additional insights into the role of efficiency on the decision whether or not to adopt organic farming. In addition, in the case of panel data, the extension of the latent class modeling to account for intra-class heterogeneity (fixed or random effects) is another interesting avenue for future research. In the case of the classic stochastic frontier, few models have been developed to deal with fixed effects (Chen et al., 2014; Wang & Ho, 2010). These models could be extended to the LCSFM to account for intra-class heterogeneity even though this adds more complexity to the likelihood function. Finally, dynamic aspects in light of adjustment costs are another future avenue of research (Minviel & Sipiläinen, 2021).

NOTES

1 As underlined in W. Greene (2010), other sample selection specifications have been developed in the literature (Kumbhakar et al., 2009, Lai, 2015). These specifications have not been considered in this article due to the substantive complexity associated with them, particularly when extending them to the LCSFM.

2 A drawback of the formulation in (2) is that the inefficiency term is assumed to be homoscedastic. This model provides biased coefficients and efficiency scores if inefficiency is heteroscedastic.

3 Throughout the manuscript, we have used f() to define various density functions.

4 The closed form of the integral in (5) can be found in Aigner et al. (1977), Kumbhakar and Lovell (2000) and estimated using maximum likelihood.

5 Equation (8) and the following equations apply to the sub-sample of AES adopters. For the sub-sample of non-adopters, we simply reverse the process where y_{it} is only observed for non-adopters.

6 In our case, only two classes are considered since the model failed to converge for three and more classes.

7 The BACON algorithm (Billor et al., 2000), in addition to graphical visualizations, has been used a priori to screen out outlying observations.

8 All monetary values were deflated by output or input price indices provided by the French statistical office (INSEE) with base year 2015.

9 "The livestock unit [...] is a reference unit which facilitates the aggregation of livestock from various species and age as per convention, via the use of specific coefficients established initially on the basis of the nutritional or feed requirement of

each type of animal [...]." (http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Livestock_unit_(LSU) accessed on December 14, 2020).

[10] Intermediate inputs comprise all the costs related to fertilizers, seeds, pesticides, feed purchases, veterinary services and products, energy consumption and other materials and services.

[11] "In areas designated as 'less-favored', agricultural production or activity is more difficult because of natural handicaps such as difficult climatic conditions, steep slopes in mountain areas or low soil productivity" (https://ec.europa.eu/jrc/en/news/classifying-areas-natural-handicaps-agricultural-aid-7294) accessed on December 14, 2020).

[12] We use capital letters when inputs and outputs are at levels: $y_1 = \ln Y_1$, $y_2 = \ln Y_2$, $x_k = \ln X_k$ $(k = 1, \ldots, 5)$.

[13] During the period studied, in France, most AES were not AES associated with being organic farming. Hence, farmers engaged in AES for organic conversion are assumed to have similar milk prices as non-organic farms. In other words, we believe that the potential issue of reverse causality associated with milk price is negligible.

[14] All computations were carried out using R software (R Core Team, 2020). The stochastic frontier models without sample selection were estimated using the package sfaR (Dakpo, Desjeux, et al., 2021). For replication purposes, all the codes used for the analysis of this article can be found in the online appendix.

[15] We thank one anonymous Reviewer for suggesting this correction.

ACKNOWLEDGMENTS

We would like to thank the Associate Editor Arne Henningsen and three anonymous rewiewers for their valuable comments. This paper is part of the LIFT ('Low-Input Farming and Territories – Integrating knowledge for improving ecosystem-based farming') project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 770747. The authors also acknowledge financial support from the Centre for Studies and Strategic Foresight (CEP) in the French Ministry of Agriculture under the framework of the research project COMPANI ('Competitiveness of the livestock sectors in France'). Remote access to the French FADN data, provided by the Ministry of Agriculture and used in this work, has been made possible within a secure environment offered by CASD ('Centre d'accès sécurisé aux données', Ref. 10.34724/CASD).

Open access funding provided by Eidgenossische Technische Hochschule Zurich.

REFERENCES

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Alvarez, A., & Arias, C. (2015). Effects of switching between production systems in dairy farming. *Bio-Based and Applied Economics*, *4*, 1–16.

Alvarez, A., & del Corral, J. (2010). Identifying different technologies using a latent class model: Extensive versus intensive dairy farms. *European Review of Agricultural Economics*, *37*, 231–250.

Alvarez, A., del Corral, J., & Tauer, L. W. (2012). Modeling unobserved heterogeneity in New York dairy farms: One-stage versus two-stage models. *Agricultural and Resource Economics Review*, *41*, 275–285.

Arata, L., & Sckokai, P. (2016). The impact of agri-environmental schemes on farm performance in five E.U. member states: A DID-matching approach. *Land Economics*, *92*, 167–186.

Barreiro-Hurlé, J., Espinosa-Goded, M., & Dupraz, P. (2010). Does intensity of change matter? Factors affecting adoption of agri-environmental schemes in Spain. *Journal of Environmental Planning and Management*, *53*, 891–905.

Battese, G. E., & Rao, D. P. (2002). Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*, *1*, 87–93.

Battese, G. E., Rao, D. S. P., & O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, *21*, 91–103.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: Blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, *34*, 279–298.

Bonfiglio, A., Henke, R., Pierangeli, F., & D'Andrea, M. R. P. (2020). Effects of redistributing policy support on farmers' technical efficiency. *Agricultural Economics*, *51*, 305–320.

Bostian, A. A., Bostian, M. B., Laukkanen, M., & Simola, A. (2019). Assessing the productivity consequences of agri-environmental practices when adoption is endogenous. *Journal of Productivity Analysis*, *53*, 141–162.

Bradfield, T., Butler, R., Dillon, E., Hennessy, T., & Kilgarriff, P. (2021). The effect of land fragmentation on the technical inefficiency of dairy farms. *Journal of Agricultural Economics*, *72*, 486–499.

Bravo-Ureta, B. E., González-Flores, M., Greene, W., & Solís, D. (2020). Technology and technical efficiency change: Evidence from a difference in differences selectivity corrected stochastic production frontier model. *American Journal of Agricultural Economics*, n/a.*103*(1), 362–385.

Bravo-Ureta, B. E., Greene, W., & Solis, D. (2012). Technical efficiency analysis correcting for biases from observed and unobserved variables: An application to a natural resource management project. *Empirical Economics*, *43*, 55–72.

Cabrera, V. E., Solis, D., & del Corral, J. (2010). Determinants of technical efficiency among dairy farms in Wisconsin. *Journal of Dairy Science*, *93*, 387–393.

Caudill, S. B., & Ford, J. M. (1993). Biases in frontier estimation due to heteroscedasticity. *Economics Letters*, *41*, 17–20.

Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics*, *13*, 105–111.

Chabé-Ferret, S., & Subervie, J. (2013). How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching. *Journal of Environmental Economics and Management*, *65*, 12–27.

Chen, Y. Y., Schmidt, P., & Wang, H. J. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics*, *181*, 65–76.

Cillero, M. M., Thorne, F., Wallace, M., & Breen, J. (2019). Technology heterogeneity and policy change in farm-level efficiency analysis: An application to the Irish beef sector. *European Review of Agricultural Economics*, *46*, 193–214.

Coyne, L., Kendall, H., Hansda, R., Reed, M. S., & Williams, D. J. L. (2021). Identifying economic and societal drivers of engagement in agri-environmental schemes for English dairy producers. *Land Use Policy*, *101*, 105174.1–15.

Dakpo, K. H., Desjeux, Y., & Latruffe, L. (2021). sfaR: Stochastic Frontier Analysis using R. CRAN R.

Dakpo, K. H., Latruffe, L., Desjeux, Y., & Jeanneaux, P. (2021). Latent class modelling for a robust assessment of productivity: Application to french grazing livestock farms. *Journal of Agricultural Economics*, *72*, 760–781.

Defrancesco, E., Gatto, P., Runge, F., & Trestini, S. (2008). Factors affecting farmers' participation in agri-environmental measures: A Northern Italian perspective. *Journal of Agricultural Economics*, *59*, 114–131.

del Corral, J., Perez, J. A., & Roibas, D. (2011). The impact of land fragmentation on milk production. *Journal of Dairy Science*, *94*, 517–525.

Dong, F., Hennessy, D. A., Jensen, H. H., & Volpe, R. J. (2016). Technical efficiency, herd size, and exit intentions in U.S. dairy farms. *Agricultural Economics*, *47*, 533–545.

Espinosa-Goded, M., Barreiro-Hurlé, J., & Dupraz, P. (2013). Identifying additional barriers in the adoption of agri-environmental schemes: The role of fixed costs. *Land Use Policy*, *31*, 526–535.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, *126*, 269–303.

Greene, W. (2010). A stochastic frontier model with correction for sample selection. *Journal of Productivity Analysis*, *34*, 15–24.

Greene, W. H. (2001). Fixed and random effects in nonlinear models, *NYU Working Paper No. EC-01-01*.

Greene, W. H. (2002). Alternative panel data estimators for stochastic frontier models. *Current Developments in Productivity and Efficiency Measurement*, University of Georgia.

Grovermann, C., Quiédeville, S., Muller, A., Leiber, F., Stolze, M., & Moakes, S. (2021). Does organic certification make economic sense for dairy farmers in Europe?–A latent class counterfactual analysis. *Agricultural Economics*, 1–12.

Hayami, Y., & Ruttan, V. W. (1970). Agricultural productivity differences among countries. *The American Economic Review*, *60*, 895–911.

Heckman, J. (1976). Discrete, qualitative and limited dependent variables. *Annals of Economic and Social Measurement*, *4*, 475–492.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Henningsen, A., & Henning, C. H. C. A. (2009). Imposing regional monotonicity on translog stochastic production frontiers with a simple three-step procedure. *Journal of Productivity Analysis*, *32*, 217–229.

Henningsen, A., Mpeta, D. F., Adem, A. S., Kuzilwa, J. A., & Czekaj, T. G. (2015). A meta-frontier approach for causal inference in productivity analysis: The effect of contract farming on sunflower productivity in Tanzania.

Jondrow, J., Lovell, C. A. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, *19*, 233–238.

Kellermann, M., & Salhofer, K. (2014). Dairy farming on permanent grassland: Can it keep up?, *Journal of Dairy Science*, *97*, 6196–6210.

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge University Press.

Kumbhakar, S. C., Tsionas, E. G., & Sipilainen, T. (2009). Joint estimation of technology choice and technical efficiency: An application to organic and conventional dairy farming. *Journal of Productivity Analysis*, *31*, 151–161.

Lai, H. P. (2015). Maximum likelihood estimation of the stochastic frontier model with endogenous switching or sample selection. *Journal of Productivity Analysis*, *43*, 105–117.

Latruffe, L., Bravo-Ureta, B. E., Carpentier, A., Desjeux, Y., & Moreira, V. H. (2017). Subsidies and technical efficiency in agriculture: Evidence from European dairy farms. *American Journal of Agricultural Economics*, *99*, 783–799.

Latruffe, L., & Nauges, C. (2014). Technical efficiency and conversion to organic farming: The case of France. *European Review of Agricultural Economics*, *41*, 227–253.

Lefebvre, M., Midler, E., & Bontems, P. (2020). Adoption of environment-friendly agricultural practices with background risk: Experimental evidence. *Environmental and Resource Economics*, *76*, 405–428.

Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. CAMBRIDGE University Press.

Mayen, C. D., Balagtas, J. V., & Alexander, C. E. (2010). Technology adoption and technical efficiency: Organic and conventional dairy farms in the United States. *American Journal of Agricultural Economics*, *92*, 181–195.

Meeusen, W., & Vandenbroeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, *18*, 435–445.

Minviel, J. J., & Latruffe, L. (2017). Effect of public subsidies on farm technical efficiency: A meta-analysis of empirical results. *Applied Economics*, *49*, 213–226.

Minviel, J. J., & Sipiläinen, T. (2021). A dynamic stochastic frontier approach with persistent and transient inefficiency and unobserved heterogeneity. *Agricultural Economics*, *52*, 575–589.

Murphy, K. M., & Topel, R. H. (2002). Estimation and inference in two-step econometric models. *Journal of Business & Economic Statistics*, *20*, 88–97.

Orea, L., & Kumbhakar, S. C. (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, *29*, 169–183.

Orea, L., Perez, J. A., & Roibas, D. (2015). Evaluating the double effect of land fragmentation on technology choice and dairy farm productivity: A latent class model approach. *Land Use Policy*, *45*, 189–198.

Parmeter, C. F., & Kumbhakar, S. (2014). Efficiency analysis: A primer on recent advances. *Foundations and Trends® in Econometrics*, *7*, 191–385.

Pérez-Méndez, J. A., Roibás, D., & Wall, A. (2020). Somatic cell counts, reproduction indicators, and technical efficiency in milk production: A stochastic frontier analysis for Spanish dairy farms. *Journal of Dairy Science*, *103*, 7141–7154.

Primdahl, J., Vesterager, J. P., Finn, J. A., Vlahos, G., Kristensen, L., & Vejre, H. (2010). Current use of impact models for agri-environment schemes and potential for improvements of policy design and assessment. *Journal of Environmental Management*, *91*, 1245–1254.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review*, *32*, 715–723.
Sauer, J., & Paul, C. J. M. (2013). The empirical identification of heterogeneous technologies and technical change. *Applied Economics*, *45*, 1461–1479.
Siebert, R., Toogood, M., & Knierim, A. (2006). Factors affecting european farmers' participation in biodiversity policies. *Sociologia Ruralis*, *46*, 318–340.
Sipiläinen, T., & Oude Lansink, A. (2005). Learning in switching to organic farming. *NJF Report*, *1*, 169–172.
Skevas, I., Emvalomatis, G., & Brummer, B. (2018). Heterogeneity of long-run technical efficiency of german dairy farms: A bayesian approach. *Journal of Agricultural Economics*, *69*, 58–75.
Solís, D., Bravo-Ureta, B. E., & Quiroga, R. E. (2007). Soil conservation and technical efficiency among hillside farmers in Central America: A switching regression model*. *Australian Journal of Agricultural and Resource Economics*, *51*, 491–510.
Vanslembrouck, I., Van Huylenbroeck, G., & Verbeke, W. (2002). Determinants of the Willingness of Belgian farmers to participate in agri-environmental measures. *Journal of Agricultural Economics*, *53*, 489–511.
Villano, R., Bravo-Ureta, B., Solis, D., & Fleming, E. (2015). Modern rice technologies and productivity in the Philippines: Disentangling technology from managerial gaps. *Journal of Agricultural Economics*, *66*, 129–154.
Wang, H. J., & Ho, C. W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics*, *157*, 286–296.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Dakpo, K. H., Latruffe, L., Desjeux, Y., & Jeanneaux, P. (2022). Modeling heterogeneous technologies in the presence of sample selection: The case of dairy farms and the adoption of agri-environmental schemes in France. *Agricultural Economics*, *53,* 422–438.
https://doi.org/10.1111/agec.12683