European Review of Agricultural Economics Vol **37 (2)** (2010) pp. 231–250
doi:10.1093/erae/jbq015
Advance Access Publication 14 May 2010

Identifying different technologies using a latent class model: extensive versus intensive dairy farms

Antonio Alvarez*,† and Julio del Corral‡
†University of Oviedo, Spain; ‡University of Castilla-La Mancha, Spain

Received May 2007; final version accepted February 2010

Review coordinated by Paolo Sckokai

Abstract

In this article, we use a stochastic frontier latent class model to estimate the technology of dairy farms according to their degree of intensification. The results are compared with a model which assumes that the technology is common to all farms. The empirical analysis uses data on a balanced panel of 130 Spanish dairy farms over the period 1999–2006. We find that the intensive technology is more productive than the extensive one and that intensive farms are more technically efficient than extensive farms.

Keywords: dairy farms, intensification, latent class model, stochastic frontier, technical efficiency, productivity

JEL classification: Q12, C23

1. Introduction

In the European Union (EU), there has been a large reduction in the number of dairy farms in recent decades.[1] As a consequence of the existing quota system, farm size has been steadily increasing. This increase in size experienced by many dairy farms has been accomplished by a change in production techniques, with many farms adopting more intensive systems. In general, intensification of production takes the form of an increase in the number of dairy cows per hectare (higher stocking rate), the acquisition of genetically improved dairy cattle and the increase of the share of concentrates in the diet.

*Corresponding author: Departamento de Economía, Universidad de Oviedo, Avda. del Cristo s/n, 33006 Oviedo, Spain. E-mail: alvarez@uniovi.es

1 For example, in the period 1990–2004, the percentage reduction in the number of dairy farms was the following: UK (38), Germany (61), Spain (73) (Source: Eurostat).

In this paper, we are concerned with the effects of intensification on the technical efficiency (TE) and productivity of dairy farms.[2] To identify extensive and intensive farms, we will assume that the technology of these groups may be different. This runs contrary to the common practice of estimating production functions under the assumption that the technology is the same for all producers. However, if farms use different technologies, then estimating a single technology for all farms is not appropriate because it can yield biased estimates of the technological characteristics.

Several dairy sector studies have addressed the issue of production heterogeneity. The most common approach has been to use methods that follow a two-stage process. In the first step, the sample is split into several groups based on some *a priori* information about farms, and in the second stage different functions are estimated for each group (e.g. Hoch, 1962; Newman and Matthews, 2006; Kumbhakar *et al*., 2009).

An alternative is to use models that separate the sample and estimate the technology for each group in only one stage. Latent class models (LCMs) belong to this category (Cameron and Trivedi, 2005: 611). An LCM assumes that there is a finite number of structures (classes) underlying the data. These models classify the sample into several groups and each farm can be assigned to a particular group using the estimated probabilities of class membership.

In this paper, we use an LCM to estimate the technology of two different dairy production systems according to their degree of intensification. Since we are interested in the efficiency of each group, the LCM is applied in a stochastic frontier framework (Greene, 2005b). The results are compared with a typical stochastic frontier model (Aigner *et al*., 1977), which assumes that the technology is common to all farms in the sample. The empirical analysis uses data on a balanced panel of 130 Spanish dairy farms over the 8-year period from 1999 to 2006. The results of the paper show that there are important technological differences between extensive and intensive dairy farms. Specifically, we find that the intensive technology is more productive than the extensive one and that intensive farms are more technically efficient than extensive farms.

The remainder of this article is organised as follows. First, we review the literature on stochastic frontier models which address unobserved heterogeneity. We then outline the stochastic frontier LCM and describe the data and the empirical model. This is followed by a presentation of the results and an analysis of farms' TE and productivity. The article concludes with a summary of the main findings and some policy implications.

2 This issue has been addressed in very few studies. For example, Hallam and Machado (1996) found mixed evidence since feed per cow was found to increase efficiency, while the opposite was true for cows per hectare. Cabrera *et al*. (2010) found that feed per cow increases TE.

2. Literature review

Stochastic frontier models were proposed by Aigner *et al*. (1977) and have been used extensively in the literature.[3] A stochastic frontier production function may be written as:

$$y = f(x) \cdot \exp(\varepsilon); \quad \varepsilon = v - u \tag{1}$$

where y represents output, x is a vector of inputs, $f(x)$ represents the technology and ε is a composed error term. The component v captures statistical noise and is assumed to follow a normal distribution centred at zero, while u is a non-negative term that reflects technical inefficiency and which is assumed to follow a one-sided distribution. The two components v and u are assumed to be independent of each other.

In recent years, several alternatives have been proposed in the literature to relax the restrictive assumption that all firms share the same technological parameters. Kalirajan and Obwona (1994), Tsionas (2002), Huang (2004) and Greene (2005b) have developed different versions of random coefficient models in which cross-firm heterogeneity is modelled in the form of continuous parameter variation. Kumbhakar *et al*. (2007) developed a non-parametric stochastic frontier using a local maximum likelihood approach.

Other papers classify the sample into several groups and estimate different functions for each group. Some of these use *a priori* information to classify the sample (e.g. Battese *et al*., 2004; Newman and Matthews, 2006), while others have used cluster analysis to split the sample (e.g. Maudos *et al*., 2002).

An alternative approach is to use statistical methods to determine the number of groups (and group membership) by estimating and testing the parameters of mixtures models (O'Donnell *et al*., 2008: 251). Beard *et al*. (1991) developed a stochastic function based on finite mixture distributions but only cross-sectional data were available. Gropper *et al*. (1999), Orea and Kumbhakar (2004), Tsionas and Kumbhakar (2004), O'Donnell and Griffiths (2006) and Greene (2005b) extended this to a panel data framework. The main advantage of LCMs over the alternatives mentioned previously is that tests can be implemented to choose the number of groups. Moreover, Orea and Kumbhakar (2004) argue that two-stage procedures, such as cluster analysis or splitting the sample based on *a priori* information, do not use the information contained in a given class to estimate the technology of firms that belong to other classes despite the fact that in most empirical applications this inter-class information may be quite important because firms belonging to different classes often share some common features.

Finally, Greene (2005a) proposed two alternative panel data estimators that he labelled as 'true random effects' and 'true fixed effects'. The main feature of these models is that time-invariant firm effects co-exist with inefficiency in order to avoid that the inefficiency term could be picking up firm

3 See Kumbhakar and Lovell (2000) or Greene (2008) for good overviews.

heterogeneity. The true random effects model, which assumes that there is a firm-specific random term to capture firm heterogeneity, has been used in Abdulai and Tietje (2007) and Farsi and Filippini (2008). On the other hand, the true fixed effects model assumes that the firm-specific term is a fixed parameter and is allowed to be correlated with the included variables. It was used in Greene (2004).

3. Stochastic frontier LCMs

We can write equation (1) as an LCM as follows:

$$y_{it} = f(x_{it})|_j \cdot \exp(v_{it}|_j - u_{it}|_j), \tag{2}$$

where subscript i denotes farm, t indicates time and j represents the different classes (groups). The vertical bar means that there is a different model for each class j.

Assuming that v is normally distributed and that u follows a half-normal distribution, the likelihood function (LF) for each farm i at time t for group j is (Kumbhakar and Lovell, 2000):

$$\mathrm{LF}_{ijt} = f(y_{it}|x_{it,}\beta_j, \sigma_j, \lambda_j) = \frac{\Phi(-\lambda_j \cdot \varepsilon_{it|j}/\sigma_j)}{\Phi(0)} \cdot \frac{1}{\sigma_j} \cdot \phi\left(\frac{\varepsilon_{it|j}}{\sigma_j}\right), \tag{3}$$

where $\varepsilon_{it|j} = y_{it} - \beta_j' x_{it}$, $\sigma_j = [\sigma_{uj}^2 + \sigma_{vj}^2]^{1/2}$, $\lambda_j = \sigma_{uj}/\sigma_{vj}$ and ϕ and Φ denote the standard normal density and cumulative distribution function, respectively.

The LF for farm i in group j is obtained as the product of the LFs in each period:

$$\mathrm{LF}_{ij} = \prod_{t=1}^{T} \mathrm{LF}_{ijt}. \tag{4}$$

The LF for each farm is obtained as a weighted average of its LF for each group j, using the prior probabilities of class j membership as weights:

$$\mathrm{LF}_i = \sum_{j=1}^{J} P_{ij}\mathrm{LF}_{ij}. \tag{5}$$

The prior probabilities must be between 0 and 1: $0 \leq P_{ij} \leq 1$. Furthermore, the sum of these probabilities for each farm must be 1: $\sum_j P_{ij} = 1$. In order to satisfy these two conditions, we parameterise these probabilities as a multinomial logit, that is:

$$P_{ij} = \frac{\exp(\delta_j q_i)}{\sum_{j=1}^{J} \exp(\delta_j q_i)}, \tag{6}$$

where q_i is a vector of 'separating variables' which are individual (farm) characteristics that sharpen the prior probabilities, and δ_j is a vector of parameters to be estimated. One group is chosen as reference in the multinomial logit.

The overall log LF is obtained as the sum of the individual log LFs:

$$\log \mathrm{LF} = \sum_{i=1}^{N} \log \mathrm{LF}_i = \sum_{i=1}^{N} \log \sum_{j=1}^{J} P_{ij} \prod_{t=1}^{T} \mathrm{LF}_{ijt}. \tag{7}$$

The log LF can be maximised with respect to the parameter set $\theta_j = (\beta_j, \sigma_j, \lambda_j, \delta_j)$ using conventional methods (Greene, 2005b). Furthermore, the estimated parameters can be used to estimate the posterior probabilities of class membership using the Bayes theorem:

$$P(jt/i) = \frac{P_{ij}\mathrm{LF}_{ijt}}{\sum_{j=1}^{J} P_{ij}\mathrm{LF}_{ijt}}. \tag{8}$$

Therefore, even though prior probabilities are time invariant, posterior probabilities can vary over time and therefore farms are allowed to switch between the extensive and intensive regimes.

4. Data and empirical model

The data used in the empirical analysis consist of a balanced panel of 130 Spanish dairy farms which were enrolled in a voluntary record-keeping programme over the 8-year period from 1999 to 2006.[4] Table 1 shows the average values of the main variables[5] for each year.

In general, the units considered are small-to-medium-sized family farms. The last four rows (milk per cow, milk per hectare, purchased feed per cow and cows per hectare) refer to variables which reflect the intensity of the farm system. It can be seen that intensification increases throughout the sample period. In contrast to other inputs, land usage does not increase over the sample period. This is due to the fact that the availability of land is very limited. Farm output grows due to an increase in the number of cows, with land staying almost constant, so farmers use more purchased feed per cow.

4 Using dairy farm sample based on a voluntary basis is usual in the literature. For instance, Ahmad and Bravo-Ureta (1996), Tauer (1998) and Newman and Matthews (2006) to name just a few.

5 All monetary variables are expressed in 2006 EUR. Farm expenses were deflated by using an index of prices paid by farmers, developed by the Spanish Statistical Institute.

Table 1. Evolution of the main variables in the sample (1999–2006)

	1999	2000	2001	2002	2003	2004	2005	2006
Milk (l)	252,123	277,489	306,360	340,280	339,757	362,725	380,843	392,504
Cows (units)	36	38	41	44	45	46	48	48
Feed (kg)	123,552	132,933	147,323	159,671	166,552	171,970	179,489	177,450
Farm expenses (EUR)	20,094	23,453	27,122	31,497	29,416	32,712	33,706	35,090
Land (ha)	18.99	19.03	19.34	19.41	19.52	19.96	20.29	20.47
Milk per cow (l)	6,931	7,091	7,307	7,579	7,404	7,638	7,830	8,011
Milk per hectare (l)	13,674	15,047	16,412	18,097	18,001	18,762	19,456	19,702
Feed per cow (kg)	3,350	3,382	3,507	3,540	3,614	3,613	3,679	3,634
Cows per hectare	1.96	2.12	2.23	2.37	2.41	2.43	2.47	2.45

Note: Monetary variables are expressed in 2006 EUR.

The empirical specification of the production function is translog.[6] We have chosen a flexible functional form in order to avoid imposing unnecessary *a priori* restrictions on the technologies to be estimated. The dependent variable is milk production measured in litres.[7] We have considered only one output since these farms are highly specialised (more than 90 per cent of farm income comes from dairy sales).[8] Five inputs were considered: cows (number of cows), purchased feed (kg), 'farm expenses' (includes veterinary expenses; purchased forage; expenditure on inputs used to produce forage crops, namely seeds, sprays, fertilisers; fuel; machinery depreciation), land (ha) and labour (which includes family labour and hired labour and is measured in man-equivalent units, i.e. a full-time adult male employee for one year). Each explanatory variable was divided by its geometric mean. Additionally, time dummy variables were introduced to control for factors that affect all farms in the same way but vary over time (1999 is the base year). In order to capture non-neutral technical change, we allowed for the interaction of inputs with a time trend. To control for different agro-climatic conditions, we used two sets of dummy variables: location dummies (seven areas) and a coastal dummy that takes a value of 1 if the farm is located in a county which is on the coast and 0 otherwise.

In line with the above, the equation to be estimated is:

$$\ln y_{it} = \beta_0|_j + \sum_{l=1}^{L} \beta_l|_j \ln x_{lit} + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L} \beta_{lk}|_j \ln x_{lit} \ln x_{kit} + \sum_{l=1}^{L} \beta_{lt}|_j \ln x_{lit} \cdot t + \sum_{t=2000}^{t=2006} \theta|_j D_t + \sum_{z=1}^{z=6} \varphi|_j \mathrm{DLOC}_{zi} + \delta|_j \mathrm{DCOAST}_i + v_{it}|_j - u_{it}|_j, \tag{9}$$

where t is a time trend, D_t are time dummies, DLOC_i are location dummies and DCOAST_i is the coastal dummy variable. It has been assumed that v follows a normal distribution centred at zero and u follows a half-normal distribution.

5. Estimation and results

The LCM in equation (9) was estimated by maximum likelihood using Limdep 9.0 (Greene, 2007). Due to the problems outlined in Greene (2005b) about using the likelihood ratio test to choose the number of

6 We tested the Cobb–Douglas against the translog functional form to determine whether the Cobb–Douglas was an adequate representation of the data and found conclusive evidence that it was not.

7 Output could have been also measured in value terms in order to account for possible differences in milk quality. However, since price differences are also due to non-quality factors (such as quantity of milk sold), using milk quantity as output seems more appropriate than total revenue.

8 Even though it was possible to buy quota in the market in some years of the sample period, some farms obtained quota from the national or local governments. On the other hand, we do not include subsidies (direct payments) in our analysis.

classes, we use information criteria, specifically, the AIC and SBIC.[9] The model with two groups was the preferred one according to both criteria.[10] The separating variables were 'purchased feed per cow' and the stocking rate (number of cows per hectare).[11] We used the average for each farm over the sample period. Moreover, since physical differences among regions may affect the choice of the technology, we also introduced the location dummies as separating variables (Kumbhakar *et al*., 2009).

The two continuous separating variables were significant and have the expected signs, being positive for the intensive group, which implies that a higher value of the variable increases the probability of assigning a farm into a more intensive group. In order to avoid possible endogeneity problems, the ratios 'milk per cow' and 'milk per land' were not considered.

The two groups correspond to two different production systems which we have labelled as 'extensive' and 'intensive' based on relevant descriptive statistics. In Table 2, we show the means of some representative variables for both groups. In general, more intensive farms produce more milk, have higher yielding cows, consume more purchased feed per cow, have more cows per hectare and, as a result, produce more milk per hectare.

Given that posterior probabilities are time variant, the classification of each farm as extensive or intensive can change over time. Consequently, the two groups of farms contain a different number of farms in each year. Table 3 shows the evolution of the relative composition of the groups over the sample period.

We now compare the LCM with a stochastic frontier, which we have labelled the 'pooled' frontier because it includes all observations without considering any kind of individual heterogeneity. The results of the estimation of the two models can be seen in Table A1.[12] We will pay special attention to the estimated technological characteristics, particularly the marginal products and the elasticity of scale.

Table 4 presents the marginal products of the two main inputs (cows and purchased feed) which were calculated by multiplying the elasticity of each observation times its average product. The values in the table are arithmetic means over the sample period. There are two kinds of differences to be analysed: differences across models (pooled model versus LCM) and differences across groups (extensive versus intensive). In order to make the comparison

9 These two statistics can be written as: $SBIC = -2 \cdot log\ LF(J) + m \cdot log\ (n)$ and $AIC = -2 \cdot log\ LF(J) = 2 \cdot m$, where $LF(J)$ is the value of the LF for J groups, m is the number of parameters used in the model and n is the number of observations. The preferred model will be that for which the value of the statistic is lowest.

10 The model with three groups does not achieve convergence. The SBIC and AIC statistics for the two-group model are −1.735 and −2.195, respectively, while for the one-group model the statistics are −1.519 and −1.727, respectively.

11 As pointed out by one referee, the results can be interpreted in terms of size instead of intensification since the effect of intensification may partially be confounded with size.

12 Almost all observations have positive elasticities of cows, feed and farm expenses, while a few of observations show negative elasticity for land and labour. Over 90 per cent of the observations have the right curvature.

Table 2. Characteristics of the estimated production systems (sample means)

	Extensive	Intensive
Observations	415	625
Milk (l)	263,919	376,391
Cows (units)	41	45
Land (ha)	21	19
Milk per cow (l)	6,439	8,162
Milk per hectare (l)	13,184	20,189
Milk per feed (l/kg)	2.07	2.22
Feed per cow (kg)	3,195	3,769
Cows per hectare	2.06	2.46
Posterior probability	0.83	0.90

Table 3. Farm classification according to their level of intensification

	Year								Total
Groups	1999	2000	2001	2002	2003	2004	2005	2006	
Extensive	77	80	75	78	80	82	76	77	625
Intensive	53	50	55	52	50	48	54	53	415

Table 4. Marginal products of inputs (sample means)

	LCM		Pooled model	
	Extensive	Intensive	Extensive	Intensive
Cows (l/cow)	2,503 (629)	4,725 (940)	2,392 (528)	3,644 (742)
Feed (l/kg)	0.78 (0.09)	0.75 (0.17)	0.94 (0.07)	0.91 (0.16)

Note: Standard deviations, calculated using the delta method, are given in parentheses.

across models more informative, the pooled marginal products are evaluated using the groups of the LCM. In this way, the comparison shows how the technological characteristics differ between models due only to differences in the estimated parameters (since the farms being compared are the same). There are important differences across models in the marginal products.

In relation to the group differences, the marginal product of cows increases with intensification in both models, with these differences being statistically significant in the LCM. This is an expected result since intensive farms generally have herds with a higher genetic level than extensive farms.

The other technological characteristic analysed is the scale elasticity. We calculate the scale elasticity for each observation. Table 5 contains the

Table 5. Scale elasticities

LCM		Pooled model	
Extensive	Intensive	Extensive	Intensive
1.027 (0.052)	1.094 (0.045)	0.974 (0.042)	1.033 (0.043)

Note: Standard deviations, calculated using the delta method, are given in parentheses.

mean and the standard deviation, while Figure 1 shows the kernel distributions of the scale elasticities for the groups of the LCM.

The scale elasticity is higher in the intensive group than in the extensive group in both models.[13] In order to test whether this difference is significant between the LCM groups, we use the *t*-test for differences between means.[14] The *t*-statistic gives a value of 22.11, so the null hypothesis that both groups have the same scale elasticity is soundly rejected. Likewise, we used Kolmogorov–Smirnov and Kruskal–Wallis tests in order to check whether the scale elasticities of both groups are equally distributed and whether both samples come from the same population. From the results of the tests, the null hypothesis was rejected in both cases. This result was unexpected since intensive farms are larger, and when farms grow they exhaust economies of scale. For this reason, it was expected that extensive farms would have higher scale elasticities than intensive farms. However, the groups use different technologies, and therefore it is possible that intensive farms have not yet exhausted their economies of scale.

These results help to provide a better understanding of some of the findings obtained in previous papers. For example, Alvarez and Arias (2003), using a similar data set, found that the vast majority of farms in this sector presented diseconomies of scale. This finding was at odds with the empirical fact that large farms buy quota from small farms. Since the price of milk is common to both groups of farms, the presence of diseconomies of scale implied an upward-sloping marginal cost curve, which in turn implied that the marginal value of quota had to be higher for small farms than for large farms. However, the finding in the present paper that intensive farms (which are larger than extensive farms) have higher scale elasticity than extensive farms resolves this apparent paradox.

13 The scale elasticity of intensive farms is decreasing over time, while the opposite is true for the extensive farms in the LCM.

14 The *t*-statistic is:

$$\frac{\bar{x}_i - \bar{x}_e}{\sqrt{((n_i - 1)\cdot S_i^2 + (n_e - 1)\cdot S_e^2)/(n_i + n_e - 2)\cdot(\sqrt{1/n_i + 1/n_e})}}$$

where $\bar{x}$ is the sample mean, S^2 is the variance, n is the number of observations and subscripts i and e indicate intensive and extensive groups, respectively. S_i^2 and S_e^2 were calculated using the delta method (Casella and Berger, 2002: 240).

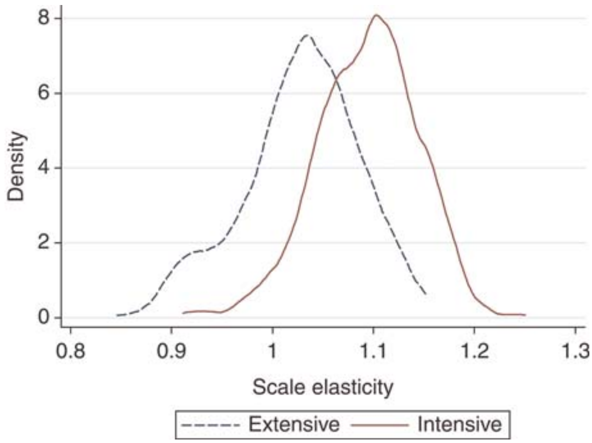

Fig. 1 Kernel distributions of the scale elasticities for the LCM groups.

6. TE and productivity

TE reflects the ability of a farm to produce the maximum level of output from a given set of inputs. In this section, we estimate farm-specific TE indexes. In a traditional stochastic frontier model (such as the frontier estimated in the 'pooled' model), the output-oriented TE index can be calculated using the following expression (the dependent variable must be in logs):

$$\mathrm{TE}_{it} = \exp(-\hat{u}_{it}), \tag{10}$$

where the inefficiency u_{it} is separated from the other error component using the formula of Jondrow *et al*. (1982).

In the stochastic frontier LCM, the calculation of TE is complicated because each farm can be assigned to several frontiers, each one with an associated probability. Two alternative methods have been proposed in the literature to calculate the TE indexes (see, for example, Orea and Kumbhakar, 2004). The first consists of measuring efficiency with respect to the most likely frontier (the one with the highest posterior probability), while the second uses a weighted average of the technical efficiencies for all the frontiers, using the posterior probabilities as weights:

$$\mathrm{TE}_i = \sum_{j=1}^{J} P\left(j/i\right) \cdot \mathrm{TE}_i|_j. \tag{11}$$

Table 6 shows the means of the TE indexes obtained using equation (11).[15]

15 Technical efficiencies obtained using the most likely frontier produce very similar results to those obtained using equation (11).

Table 6. Average TE

LCM		Pooled model	
Extensive	Intensive	Extensive	Intensive
0.931 (0.042)	0.971 (0.014)	0.865 (0.052)	0.931 (0.031)

Note: Standard deviations are given in parentheses.

An expected result is that the LCM attributes higher mean TE to farms than the pooled model in the two production systems. The reason for this is that the LCM allows each farm's TE to be measured with respect to its own (weighted) frontier. We thus avoid Stigler's critique (Stigler, 1976) that the efficiency indexes may be reflecting differences in the technology employed by the farms.[16]

With respect to differences across groups, intensive farms have higher TE than extensive farms in both models, and the difference is statistically significant.[17] This interesting result suggests that intensive farms are closer to their own frontier than extensive farms. This conclusion coincides with our perception that intensive systems are easier to manage and therefore farmers are more likely to stay close to the frontier (i.e. to make fewer mistakes). This argument is supported by the fact that, in intensive systems, feeding is based on purchased feed, while extensive systems base feeding on forage crops produced on the farm. This means that extensive farmers carry out more tasks (planting and harvesting, making silage, etc.), which may make them more likely to make mistakes. On the other hand, the fact that intensive systems base feeding on purchased feed means that they have to make fewer decisions, which suggests that these systems may be easier for farmers to manage.[18]

Next we turn to the analysis of productivity. We first test which technology is more productive and then we decompose total factor productivity (TFP) growth into technical change, efficiency change and scale change. In line with Kumbhakar *et al*. (2009), we consider that if the predicted output of technology 'a' is larger than the predicted output of technology 'b' for given input quantities, then technology 'a' is locally above technology 'b'. It is important

16 The estimated mean efficiency levels are higher than those obtained in some previous dairy farm studies (e.g. Tauer and Belbase, 1987; Kumbhakar *et al*., 1989; Ahmad and Bravo-Ureta, 1996; Alvarez *et al*., 2006) but similar to Abdulai and Tietje (2007). See Bravo-Ureta *et al*. (2007) for TE in dairy farming studies.

17 The normality of technical efficiencies was rejected using a test based on skewness and kurtosis. Therefore, the *t*-test, which is based on normality, is not valid. Hence, we used Kolmogorov–Smirnov and Kruskal–Wallis tests (Farsi and Filippini, 2008; Kopsakangas-Savolainen and Svento, 2008).

18 An extreme example of an intensive system is that of a farmer who has no land and who buys all the feed. This feed, known as 'total mixed ration' (TMR), consists of a mix of forage and concentrates. In many farms, the feed is distributed automatically. Therefore, once the technical advisor has set a particular ration, the farmer only has to make a phone call to order another truckload when the current one is almost finished. This extreme example can help to understand our claim that the intensive system is easier to manage than the extensive system.

to note that unless one technology is above the other for all data points we cannot conclude that is more productive. Table 7 shows the descriptive statistics of intensive and extensive predicted outputs for the LCM groups, while Table 8 shows the distribution of the ratio between the intensive predicted output and the extensive predicted output.

We find that the intensive frontier is located above the extensive frontier for almost 99 per cent of the observations. Likewise, the average productivity gap in favour of the intensive technology is between 9 and 18 per cent. This result is in line with other studies (Oude Lansink *et al*., 2002; Kumbhakar *et al*., 2009) which found that intensive technologies are more productive.

This result, together with the fact that the average TE (with respect to their own frontier) is higher for intensive farms than for extensive farms, allows us to conclude that, in our sample, the farms which use the intensive technology are more productive than the extensive ones.[19]

TFP growth is the difference between the rate of change of output and the rate of change of an input quantity index. A convenient way to construct an input quantity index is by using the output elasticities as weights. Thus, TFP growth can be expressed as:

$$\dot{\mathrm{TFP}}_{it} = \ln \dot{y}_{it} - \sum_{l=1}^{L} \eta_l \ln \dot{x}_{lit}, \tag{12}$$

where η_l are output elasticities for the two classes and the dot over a variable indicates a growth rate.

Taking derivatives with respect to time in equation (9), the output growth is:

$$\begin{aligned} \ln \dot{y}_{it} = & \sum_{l=1}^{L} \beta_l|_j \ln \dot{x}_{lit} + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L} \beta_{lk}|_j \ln \dot{x}_{lit} \ln \dot{x}_{kit} + \sum_{l=1}^{L} \beta_{lt}|_j \ln \dot{x}_{lit} t \\ & + \sum_{l=1}^{L} \beta_{lt}|_j \ln x_{lit} + \Delta\lambda_t|_j - \Delta u_{it}|_j. \end{aligned} \tag{13}$$

Inserting equation (13) into equation (12) and rearranging terms we obtain:

$$\begin{aligned} \dot{\mathrm{TFP}}_{it}|_j = & \left(\sum_{l=1}^{L} \beta_{lt}|_j \ln x_{lit} + \Delta\lambda_t|_j \right) + \left(-\Delta u_{it}|_j \right) \\ & + \left((\eta|_j - 1) \sum_{l=1}^{L} \frac{\eta_l|_j}{\eta|_j} \ln \dot{x}_{lit} \right), \end{aligned} \tag{14}$$

where η is the scale elasticity. The first term in parentheses is technical change, the second term represents efficiency change and the last one is the

19 It should be noted that the possible environmental effects are not taken into account in the analysis, which is based only on conventional inputs and outputs.

Table 7. Farm productivity (in litres) under different technologies

Year		Mean	Std Dev.	Min	Max
1999	$\hat{y}_{int}$	281,022	129,103	107,455	749,010
	$\hat{y}_{ext}$	255,164	105,082	110,069	634,454
	$\hat{y}_{int}/\hat{y}_{ext}$	1.09	0.08	0.85	1.35
2000	$\hat{y}_{int}$	313,051	135,968	115,317	791,221
	$\hat{y}_{ext}$	272,554	111,638	115,973	666,209
	$\hat{y}_{int}/\hat{y}_{ext}$	1.14	0.08	0.99	1.53
2001	$\hat{y}_{int}$	345,444	149,882	144,750	893,563
	$\hat{y}_{ext}$	301,607	126,557	134,631	776,770
	$\hat{y}_{int}/\hat{y}_{ext}$	1.14	0.07	1.00	1.47
2002	$\hat{y}_{int}$	380,888	172,303	138,580	1,043,870
	$\hat{y}_{ext}$	341,231	153,755	126,571	934,511
	$\hat{y}_{int}/\hat{y}_{ext}$	1.12	0.07	0.99	1.51
2003	$\hat{y}_{int}$	382,728	167,022	152,345	906,881
	$\hat{y}_{ext}$	333,912	146,607	130,734	779,452
	$\hat{y}_{int}/\hat{y}_{ext}$	1.15	0.05	1.03	1.43
2004	$\hat{y}_{int}$	408,316	176,335	147,284	896,038
	$\hat{y}_{ext}$	357,646	157,717	126,307	843,068
	$\hat{y}_{int}/\hat{y}_{ext}$	1.15	0.05	1.02	1.39
2005	$\hat{y}_{int}$	430,316	199,210	121,835	1,101,440
	$\hat{y}_{ext}$	381,107	185,894	110,606	1,076,020
	$\hat{y}_{int}/\hat{y}_{ext}$	1.14	0.06	1.01	1.35
2006	$\hat{y}_{int}$	446,664	215,694	156,404	1,109,150
	$\hat{y}_{ext}$	382,843	196,274	118,935	1,009,130
	$\hat{y}_{int}/\hat{y}_{ext}$	1.18	0.07	1.06	1.55

Table 8. Distribution of the predicted output ratios (intensive versus extensive technologies)

$\hat{y}_{int}/\hat{y}_{ext}$ interval	Number of observations	Percentage of observations
$\hat{y}_{int}/\hat{y}_{ext} \leq 1$	16	1.54
$1 < \hat{y}_{int}/\hat{y}_{ext} \leq 1.1$	243	23.37
$1.1 < \hat{y}_{int}/\hat{y}_{ext} \leq 1.2$	644	61.92
$1.2 < \hat{y}_{int}/\hat{y}_{ext} \leq 1.3$	108	10.38
$1.3 \leq \hat{y}_{int}/\hat{y}_{ext} < 1.4$	21	2.02
$\hat{y}_{int}/\hat{y}_{ext} > 1.4$	8	0.77

scale change (see Kumbhakar and Lovell, 2000, for further details). Table 9 shows the decomposition of TFP growth.

It can be seen that, in the LCM, the TFP change is larger for the intensive group (actually, the TFP change is negative for extensive farms). Several extensive farms in the LCM face decreasing returns to scale and given that

Table 9. Components of TFP growth

	LCM		Pooled model	
	Extensive	Intensive	Extensive	Intensive
Technical change	0.004	0.013	0.002	0.007
Efficiency change	−0.006	0.005	−0.016	0.008
Scale change	−0.006	0.004	0.003	0.003
TFP growth	−0.008	0.022	−0.011	0.018

farm size has increased over the sample period leads to their scale change being negative. This is not the case in the pooled model in which, even though the average scale elasticity is less than 1 in the extensive group, farms have increased size, making the scale change positive. The scale effect is positive, on the other hand, for the intensive group.

7. Discussion

In this section, we will exploit the results obtained in the previous sections in order to answer two questions: (i) How important is the assumption that the technology is homogeneous for all farms? (ii) What are the implications of the results in relation to the recent changes in the Common Agricultural Policy (CAP) of the EU?

The first question can be answered using the results obtained in the previous sections. The pooled model estimates a general technology which seems to misrepresent the technology of the different groups. In particular, the marginal product of purchased feed, which is a key technological parameter for dairy farmers, is likely to be overestimated. Since we were not able to find other estimates with which to compare our values, we looked for a different benchmark. One possibility is to compare the marginal product with the ratio of the price of purchased feed to the price of milk (these are the only prices in our database) for the two groups. The two values should be similar if farms are maximising profits since they will use variable inputs up to the point where the value of the marginal product equals the input price. The price ratio takes the following values: 0.70 (extensive) and 0.69 (intensive). A comparison of these values with the estimated marginal products (Table 4) allows us to conclude that the LCM produces estimates which are more consistent with profit-maximising behaviour, whereas the estimates of the pooled model seem to be too large in this respect.

Finally, it is interesting to analyse the implications of our findings in the context of the recent reforms of the CAP. Among various other objectives, two main concerns are behind these reforms: one is the need to enhance the competitiveness of European farms in order to compete in increasingly globalised markets; the other concern is the environmental impact of intensive agriculture.[20] However, we find that intensive farms are more productive than extensive ones. Therefore, even though the EU has the objective of promoting extensification in order to reduce the environmental problems associated with intensive systems, it seems that it will be difficult to go against the market in the sense of preventing the move towards intensification, which in our opinion is very likely to continue in the near future.

8. Conclusions

In this paper, we have identified and studied the technological differences of two groups of dairy farms based on the degree of intensification of the production system. For this purpose, we have estimated a stochastic production frontier using an LCM. The results are compared with a model which assumes that the technology is the same for all farms.

Several interesting conclusions can be drawn from our empirical analysis. First of all, from a methodological point of view, we have shown that the pooled model estimates a general technology which misrepresents the technology of the different groups. In particular, the marginal product of purchased feed, which is a very important technological parameter, appears to be overestimated.

From an agricultural policy point of view, an important result is that intensive farms have higher TE than extensive farms. Moreover, the intensive technology is more productive for a given level of inputs. These results do not mean that the intensive system should be proposed for all farms, as this decision has to be based on the criterion of profitability. However, this finding does suggest that an important part of the inefficiency detected in previous studies can be attributed to poor management of forage production.

A final remark has to do with the implications of these findings in the context of the CAP. The recent reforms of the CAP are aimed at improving environmental conditions, the latter being associated with the use of more extensive systems. However, given that our results indicate that intensive farms are more productive than extensive farms, it seems likely that the trend towards intensification will continue.

In any case, this does not mean that intensive milk production is incompatible with a reduction in the environmental problems associated with dairy farming. Intensification offers some opportunities for environmental enhancement since some of the problems associated with dairying come from insufficient control of nutrient flows and animal behaviour, and more intensive systems can give farmers more control of animals and effluent. Therefore, there is an opportunity to take advantage of the infrastructure and management skills available on more intensive farms to improve environmental performances.

20 It is noteworthy to indicate that intensive systems can generate less greenhouse gas emissions per unit of milk produced than extensive ones (Luo *et al.*, 2008).

Acknowledgements

The authors wish to thank Christine Amsler, Carlos Arias, Angel Gavilán, William Greene, Chris O'Donnell, Luis Orea, Juan Prieto, David Roibás, Loren Tauer, Alan Wall, three anonymous referees and the journal editor for very useful comments and suggestions. The authors also acknowledge the financial support of the Spanish Ministry of Science through the project SEJ2007-66979.

References

Abdulai, A. and Tietje, H. (2007). Estimating technical efficiency under unobserved heterogeneity with stochastic frontier models: application to Northern German dairy farms. *European Review of Agricultural Economics* 34: 393–416.

Ahmad, M. and Bravo-Ureta, B. (1996). Technical efficiency measures for dairy farms using panel data: a comparison of alternative model specifications. *Journal of Productivity Analysis* 7: 399–415.

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

Alvarez, A. and Arias, C. (2003). Diseconomies of size with fixed managerial ability. *American Journal of Agricultural Economics* 85: 134–142.

Alvarez, A., Arias, C. and Orea, L. (2006). Explaining differences in milk quota values: the role of economic efficiency. *American Journal of Agricultural Economics* 88: 182–193.

Battese, G., Rao, P. and O'Donnell, C. (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.

Beard, T., Caudill, S. and Gropper, D. (1991). Finite mixture estimation of multiproduct cost functions. *Review of Economics and Statistics* 73: 654–664.

Bravo-Ureta, B., Solís, D., Moreira, V. H., Maripani, J., Thiam, A. and Rivas, T. (2007). Technical efficiency in farming: a meta-regression analysis. *Journal of Productivity Analysis* 27: 57–72.

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

Cameron, C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Casella, G. and Berger, R. (2002). *Statistical Inference*. 2nd edn. Pacific Grove, CA: Duxbury Press.

Farsi, M. and Filippini, M. (2008). Effects of ownership, subsidization and teaching activities on hospital costs in Switzerland. *Health Economics* 17: 335–350.

Greene, W. (2004). Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national care systems. *Health Economics* 13: 959–980.

Greene, W. (2005a). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23: 32.

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

Greene, W. (2007). Limdep Version 9.0. New York: Econometric Software.

Greene, W. (2008). The econometric approach to efficiency analysis. In: H. Fried, C. A. K. Lovell and S. Schmidt (eds), *The Measurement of Productive Efficiency and Productivity Growth*. New York: Oxford University Press.

Gropper, D., Caudill, S. and Beard, T. (1999). Estimating multiproduct cost functions over time using a mixture of normals. *Journal of Productivity Analysis* 11: 201–218.

Hallam, D. and Machado, F. (1996). Efficiency analysis with panel data: a study of Portuguese dairy farms. *European Review of Agricultural Economics* 23: 79–93.

Hoch, I. (1962). Estimation of production function parameters combining time-series and cross-section data. *Econometrica* 30: 34–53.

Huang, H. (2004). Estimation of technical inefficiencies with heterogeneous technologies. *Journal of Productivity Analysis* 21: 277–296.

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

Kalirajan, K. P. and Obwona, M. B. (1994). Frontier production function: the stochastic coefficients approach. *Oxford Bulletin of Economics and Statistics* 56: 87–96.

Kopsakangas-Savolainen, M. and Svento, R. (2008). Estimation of cost-effectiveness of the Finnish electricity distribution utilities. *Energy Economics* 30: 212–229.

Kumbhakar, S., Biswas, B. and Bailey, D. (1989). A study of economic efficiency of Utah dairy farmers: a system approach. *Review of Economics and Statistics* 71: 595–604.

Kumbhakar, S. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.

Kumbhakar, S., Park, B., Simar, L. and Tsionas, E. (2007). Nonparametric stochastic frontiers: a local maximum likelihood approach. *Journal of Econometrics* 137: 1–27.

Kumbhakar, S., Tsionas, E. and Sipiläinen, 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.

Luo, J., Ledgard, S., de Klein, C., Lindsey, S. and Kear, M. (2008). Effects of dairy farming intensification on nitrous oxide emissions. *Plant Soil* 309: 227–237.

Maudos, J., Pastor, J. and Pérez, F. (2002). Competition and efficiency in the Spanish banking sector: the importance of specialization. *Applied Financial Economics* 12: 505–516.

Newman, C. and Matthews, A. (2006). The productivity performance of Irish dairy farms 1984–2000: a multiple output distance function approach. *Journal of Productivity Analysis* 26: 191–205.

O'Donnell, C. and Griffiths, W. (2006). Estimating state-contingent production frontiers. *American Journal of Agricultural Economics* 88: 249–266.

O'Donnell, C., Rao, P. and Battese, G. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics* 34: 231–255.

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

Oude Lansink, A., Pietola, K. and Bäckman, S. (2002). Efficiency and productivity of conventional and organic farms in Finland 1994–1997. *European Review of Agricultural Economics* 29: 51–65.

Stigler, G. (1976). The Xistence of X-efficiency. *American Economic Review* 66: 213–216.

Tauer, L. (1998). Productivity of New York dairy farms measured by nonparametric Malquist indices. *Journal of Agricultural Economics* 49: 234–249.

Tauer, L. and Belbase, K. (1987). Technical efficiency of New York dairy farms. *Northeastern Journal of Agricultural and Resource Economics* 16: 10–16.

Tsionas, E. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics* 17: 127–147.

Tsionas, E. and Kumbhakar, S. (2004). Markov switching stochastic frontier model. *Econometrics Journal* 7: 398–425.

Appendix A

Table A1. Estimation of the latent class and pooled models

	'Pooled' stochastic frontier		Extensive group		Intensive group	
	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio
Frontier						
Constant	12.600***	541	12.441***	442	12.605***	499
Cows	0.449***	8.1	0.431***	7.5	0.701***	11.5
Feed	0.433***	12.2	0.282***	6.3	0.310***	8.3
Farm$	0.113***	5.5	0.202***	9.5	0.071***	3.5
Land	0.002	0.1	−0.017	−0.5	0.064**	2.1
Labour	0.028	1.0	0.059*	1.8	−0.025	−0.9
0.5 · Cows · Cows	−1.189***	−5.2	−0.756***	−3.0	0.066	0.2
0.5 · Feed · Feed	−0.194**	−2.6	−0.029	−0.3	0.026	0.3
0.5 · Farm$ · Farm$[a]	−0.003	−0.1	0.084***	3.8	−0.028	−0.9
0.5 · Land · Land	−0.238***	−3.5	0.035	0.4	−0.080	−1.0
0.5 · Labour · Labour	−0.008	−0.5	0.018	1.3	0.019	0.3
Cows · Feed	0.483***	4.5	0.273**	2.2	−0.080	−0.5
Cows · Farm$	0.247***	4.3	0.220***	3.1	0.146*	1.9
Cows · Land	0.156*	1.8	−0.082	−1.0	0.065	0.5
Cows · Labour	0.129*	1.7	0.317***	5.0	−0.126	−1.3
Feed · Farm$	−0.144***	−3.7	−0.167***	−3.4	−0.070*	−1.7
Feed · Land	−0.060	−1.0	0.009	0.1	−0.131	−1.6
Feed · Labour	−0.057	−1.1	−0.078	−1.3	0.122**	2.1
Farm$ · Land	0.029	0.9	−0.029	−1.0	0.006	0.2
Farm$ · Labour	0.026	0.9	−0.135***	−4.7	0.022	0.7
Land · Labour	0.035	0.8	0.041	0.9	0.043	0.9
D2000	0.010	0.8	−0.011	−0.8	0.025**	2.2
D2001	0.013	1.0	−0.004	−0.2	0.030***	2.6
D2002	0.034**	2.5	0.036**	2.3	0.043***	3.6
D2003	0.009	0.7	−0.009	−0.6	0.029**	2.4
D2004	0.025*	1.8	0.012	0.7	0.047***	3.9
D2005	0.049***	3.7	0.033*	1.9	0.065***	5.2

(*continued*)

Table A1. Estimation of the latent class and pooled models (*continued*)

	'Pooled' stochastic frontier		Extensive group		Intensive group	
	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio	Coefficient	*t*-Ratio
D2006	0.067***	4.9	0.025	1.4	0.090***	7.0
DLOC1	0.059***	3.6	0.134***	6.3	0.054***	3.7
DLOC2	0.074***	6.7	0.105***	7.2	0.072***	7.0
DLOC3	0.087***	5.0	0.104***	4.8	0.063***	4.0
DLOC4	0.066***	2.9	0.136***	5.2	0.082***	3.8
DLOC5	0.115***	7.8	0.171***	10.1	0.085***	6.1
DLOC6	0.115***	4.2	0.226***	7.6	0.148***	4.3
Cows · trend	−0.011	−1.0	0.002	0.2	−0.028**	−2.3
Feed · trend	−0.003	−0.4	0.016*	1.9	0.005	0.7
Farm$ · trend	0.011***	2.7	−0.008*	−1.9	0.008**	2.0
Land · trend	0.002	0.4	0.014**	2.1	0.005	0.9
Labour · trend	−0.003	−0.5	−0.007	−1.1	0.002	0.3
DCOAST	0.022	1.3	0.047**	2.3	0.032*	1.9
Probabilities						
Constant			–		−4.977***	−3.4
Cows per hectare			–		0.001***	3.2
Feed per cow			–		0.77**	2.0
DLOC1			–		−0.836	−0.8
DLOC2			–		−0.132	−0.2
DLOC3			–		0.376	0.4
DLOC4			–		−0.064	−0.1
DLOC5			–		−0.145	−0.2
DLOC6			–		−1.826	−1.2
$\lambda = \sigma_u/\sigma_v$	1.880***	14	4.582***	3.3	1.514***	4.7
$\sigma = [\sigma_v^2 + \sigma_u^2]^{1/2}$	0.144***	1,200	0.13***	17	0.096***	15
Log LF	905		1,190			

*Significance at the 10 per cent level.
**Significance at the 5 per cent level.
***Significance at the 1 per cent level.
[a]Farm$ stays for farm expenses.