Gross output and livestock sales modelling in Spanish extensive farms using PLSR

P. Gaspar a*, F.J. Mesías b, M. Escribano a and F. Pulido b

(a) Departamento de Producción Animal y Ciencia de los Alimentos. Universidad de Extremadura

(b) Departamento de Economía. Universidad de Extremadura.

Contact: Paula Gaspar. Departamento de Producción Animal y Ciencia de los Alimentos.

Paper prepared for presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies". Sevilla, Spain, January 29th -February 1st, 2008

Copyright 2007 by P. Gaspar, F.J. Mesías, M. Escribano and F. Pulido. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Abstract

The aim of this paper is to model some production variables in extensive livestock farms located in the dehesa ecosystem. We intend to use not only purely economic variables in the construction of the model, but also structural variables in order to identify the characteristics of the farms that have the higher influence. Another objective is to be able to predict these variables at the farm level, using structural variables that are easy to measure. The data used in this work were obtained from a questionnaire survey to the holders/managers of a sample of 69 dehesa farms in Extremadura (SW Spain). The statistical methodology used for the construction of the model was Partial Least Square Regression (PLSR). It can be concluded that the variables relative to farm intensification, to labour and especially to Iberian pig breeding, are those that take part mainly in the model.

Key words: dehesa, livestock farming systems, partial least square regression, gross output.

1. Introduction

The dehesa is an agroforestry system used for livestock range farming characterized by its mix of pasture and evergreen oak stands. It originated from the traditional Mediterranean forest, and indeed human intervention has been fundamental in maintaining the dehesa ecosystem as such, because the use of appropriate cultural practices has conserved the tree stratum, thus avoiding scrub invasion and increasing the system's efficiency (Coelho, 1994; Escriabano and Pulido, 1998). Mixed-species grazing of beef cattle, sheep, and Iberian pigs is often practised to more efficiently utilize grazing resources. The orientation is to meat production and the sale of animals for intensive fattening (Pulido et al., 1999).

These rangelands constitute the most representative extensive livestock farming system of the Iberian Peninsula. It is localized in the SW quadrant, occupying a total surface area of 5.8 million ha in Spain and 0.5 million ha in Portugal (Joffre et al., 1999). The Spanish region of Extremadura is the main area, with 2.2 million hectares of dehesas.

Livestock farming systems in dehesas are based on their productive diversity and therefore the use of different livestock species simultaneously is frequent for a better use of the different resources. The average size of the farms is around 500ha (Escribano et al. 2001a; Milán et al. 2006; Plieninger and Wilbrand, 2001; Plieninger et al., 2004; Porras et al., 2000) and the average level of stocking rate is 0.37 UGM/ha (Escribano et al., 2002), lower than other European systems considered as extensive (Lasseur, 2005; Milán et al., 2003; Serrano et al., 2004).

The main products obtained from dehesa systems are meat and live animals of the different species, most of them, considered as high quality products. The importance of these livestock farming systems is firstly for their contribution to the regional economy and secondly because the persistence of these types of farms guarantees the maintenance of this complex and particularly sensitive ecosystem.
Due to the complexity of these systems the purpose of this paper is to model some production variables in extensive livestock farms located in the dehesas. We intend to use not only purely economic variables in the construction of the model, but also structural variables in order to identify the characteristics of the farms that have the higher influence. Another objective is to be able to predict these variables at the farm level, using structural variables that are easy to measure.

2. Material and methods

The data used were obtained from survey questionnaires conducted with dehesa farm owners or managers in the Region of Extremadura (SW Spain). The surveys were carried out in 2004 and 2005 as part of the INTERREG-III project "Development of an information system for the environmental and economic management of the dehesa/montado ecosystem in Extremadura and Alentejo".

2.1. Sampling and questionnaire design

The surveyed farms were representative of dehesas in Extremadura, and were selected randomly according to forestry, soil-type, livestock, and economic size criteria. The sample consisted of 69 farms. The questionnaire comprised to principal blocks – a technical part to gather descriptive data on the area, infrastructure, and livestock management regime of the farm, and an economic part to collect data on the intermediate consumption and output generated in the system.

The surveys were conducted on site, and the interviewee was generally the farm's owner or manager. The interviews were carried out in each farm twice – once in 2004 for the data of the 2003 financial year, and again in 2005 for the 2004 financial year. The indicators used in the study are the mean values of these two years.

2.2 Methodological criteria and creation of the indicators

The technical indicators were designed on the basis of the work of Escribano et al. (2001b and 2002) for the stocking rate, Escribano et al. (2001a) for land uses, and Martin et al. (2001) for livestock productivity. The method used for the design of the economic indicators was a microeconomic adaptation of the Economic Accounts of Agriculture and Forestry (European Communities, 2000) with some methodological changes that have been implemented through several research works aimed at providing a rigorous measure of the economic resources of dehesa farms (Campos, 1993; Pulido and Escribano, 1994; Pulido, 2003). The variables used in this study, their measurement units, and the codes with which they are referred to throughout the paper, are listed in tables 1 and 2.

Table 1. Production indicators, units and codes (dependent variables)

<table>
<thead>
<tr>
<th>Production Indicators</th>
<th>Unit</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Livestock sales/ha UAA</td>
<td>€/ha</td>
<td>LvSales</td>
</tr>
<tr>
<td>Gross output/ha UAA</td>
<td>€/ha</td>
<td>GO</td>
</tr>
</tbody>
</table>
Table 2. Technical indicators, units and codes (independent variables)

<table>
<thead>
<tr>
<th>Technical Indicators</th>
<th>Unit</th>
<th>Code</th>
<th>Technical Indicators</th>
<th>Unit</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cattle stocking rate</td>
<td>LU/ha</td>
<td>C</td>
<td>Rented UAA/Total UAA</td>
<td>%</td>
<td>REN</td>
</tr>
<tr>
<td>Sheep stocking rate</td>
<td>LU/ha</td>
<td>S</td>
<td>Wooded UAA/Total UAA</td>
<td>%</td>
<td>WOO</td>
</tr>
<tr>
<td>Goat stocking rate</td>
<td>LU/ha</td>
<td>G</td>
<td>Scrubland UAA/Total UAA</td>
<td>%</td>
<td>SCRUB</td>
</tr>
<tr>
<td>Pig stocking rate</td>
<td>LU/ha</td>
<td>P</td>
<td>Pasture UAA/Total UAA</td>
<td>%</td>
<td>PAST</td>
</tr>
<tr>
<td>Total stocking rate</td>
<td>LU/ha</td>
<td>SR</td>
<td>Irrigated UAA/Total UAA</td>
<td>%</td>
<td>IRR</td>
</tr>
<tr>
<td>Permanent AWU/100 ha UAA</td>
<td>AWU/100ha</td>
<td>PER</td>
<td>Cultivated UAA/Total UAA</td>
<td>%</td>
<td>CUL</td>
</tr>
<tr>
<td>Temporary AWU/100 ha UAA</td>
<td>AWU/100ha</td>
<td>TEM</td>
<td>Reforested UAA/Total UAA</td>
<td>%</td>
<td>FOR</td>
</tr>
<tr>
<td>Family AWU/100 ha UAA</td>
<td>AWU/100ha</td>
<td>FAM</td>
<td>Livestock Subsidies</td>
<td>€/ha</td>
<td>LS</td>
</tr>
</tbody>
</table>

The selection of the technical indicators has been carried out taking into account, on the one hand, the fact that they were easy to obtain in a survey to producers, and on the other hand that they were significant enough to provide a good model for economic prediction.

2.3. Statistical Analysis

Partial least square regression (PLSR) (Martens and Næs, 1993; Esbesen, 2002) was carried out with the Unscrambler software (v. 9.2) (Camo AS, Oslo, Norway). PLSR is a multivariate calibration method, by which two sets of data, X (e.g. technical variables of the farms) and Y (e.g. production indicators) are related by means of regression. The purpose of PLSR is to establish a linear model, which enables the prediction of Y from the measured X, using an equation of the form

\[ t_i = b_0 + \sum_{j=1}^{m} b_j x_{ij} \]

where \( t_i \) was the value of the economic indicator for the ith farms, \( b_0 \) was the \( y \)-intercept, and \( b_j \) was the regression coefficient for the j-th prediction parameters (\( X_j \)) in the model. The contribution of each variable to predict the economic variables was evaluated using the regression coefficients obtained for the standardised variables. These coefficients allow the selection of those variables that most contributed to the prediction. Martens’s uncertainty test (Esbesen, 2002) was used to eliminate noisy variables. The model was validated using full cross validation (‘‘leave one out’’), and only validated results are reported.

PLSR have been widely applied in agricultural research as an alternative to multiple linear regression and principal components regression (Poveda et al., 2004; Ruiz et al., 2002; Downey et al., 2005; Thybo et al., 2003). In livestock management research, where the relationships among the variables are complex and the number of observations is usually small, the use of PLSR as opposed to PCR shows numerous advantages, as the possibility of modelling several variables at the same time and. Another advantage is no need for the variables to follow a normal distribution, that allowing good...
adjustments when the number of variables is high compared with the number of cases (Rougoor et al., 2000).

### 3. Results

The resulting model has used 2 Principal Components (PC). The first PC explains the 41.26% of the variation of the dependent variables, while the second explains 23.26%. The loadings of the dependent and independent variables for these two PCs are shown in figure 1.

---

**Figure 1.** Loading of PC1 and PC2 obtained from partial least squares regression analysis (PLSR) of technical indicators (independent variables) and production indicators (dependent variables).

**Figure 2.** Positioning of the 69 farms according to scores obtained for PCs 1 and 2 and typified according to livestock species and their stocking rates.

---

The first component can be defined as a factor of intensification, as the technical variables with the highest correlation are stocking rate and livestock subsidies. It must be highlighted that all the production indicators are located in the positive axis of the PC1, which implies that the farms located in that area will be the ones with higher productions.

The second component is related to Iberian pig orientation. Variables highly and positively correlated are pig stocking rate and wooded area, while those negatively correlated are the ones more important for non pigment farms: livestock subsidies, family labour and total stocking rate.

Each farm gets a score for every PC, which allow drawing them in a plot. Figure 2 shows the position of the farms according to their stocking rates and to the main livestock species bred in the farm. One can see from figure 2 a clear positioning of the farms depending on their stocking rates. Those with high stocking rates (more intensified) are located basically in the positive area of the PC1, which
implies higher economic indicators. The farms with medium and low stocking rates appear as we move throughout the X axis.

Livestock species bred in the farm give also interesting information, especially concerning Iberian pigs. This appears in figure 2 and it can be seen that most of the farms with Iberian pigs are more profitable than those that do not breed them. This is due both to the good complementariness of extensive pigs with the other livestock in the dehesa (cattle and sheep) and also to the good market trends for Iberian pig products during the last few years. Both reasons affect positively the farms in which Iberian pigs are bred.

The resulting prediction equations obtained for the two production variables appear in table 3. Each one of the two production indicators is explained by 8 technical indicators which finally were significant.

Table 3. Prediction equations

<table>
<thead>
<tr>
<th>Equation</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{LvSales} = 28.59 + 1.930P - 49.56\text{RE}N + 52.4\text{WOO} - 43.87\text{PAST} + 197.97\text{SR} + 312.95\text{TEM} + 51.72\text{FAM} + 0.56\text{LS} )</td>
<td></td>
</tr>
<tr>
<td>( \text{GO} = 193.79 + 2.058P - 49.52\text{RE}N + 55.71\text{WOO} - 39.70\text{PAST} + 327.56\text{SR} + 404.66\text{TEM} + 93.61\text{FAM} + 1\text{LS} )</td>
<td></td>
</tr>
</tbody>
</table>

The model’s statistics for livestock sales were: correlation coefficient in the validation stage, 0.84, explained variance 70% and RMSEP 76.50 €/ha; for gross product the correlation coefficient was 0.85, the explained variance, 72% and the RMSEP, 97.36 €/ha. Figures 3 and 4 show the observed and predicted values for each variable.

These results are comparable with other works about extensive livestock systems modelling, although we have used multiple linear regression calibration methods. Pérez et al. (2001) analyzed 49 sheep farms and obtained production and benefit functions using technical and economical variables. The final functions explained around 84% of the variances using between five (production function) and eight variables (enterprise benefit function). Acero (2002), in a study of 63 goat farms, developed
models of unitary costs, production and net result. The explained variance of the final models varied from 71% to 93%.

4. Conclusions

It can be said that the variables relative to farm intensification, to labour and especially to Iberian pig breeding, are those that take part mainly in the model. The model developed can predict, with an acceptable error and using structural variables that are easy to measure, two very important variables for the farmers, as gross product and livestock sales. For that reason, it has a great interest as a tool to help the managers in this type of farms.

References


