

Estimativa dos Coeficientes de Afetação dos Custos às Atividades Agrícolas Usando a Teoria da Máxima Entropia¹

Maria do Socorro Rosário², Rui Manuel de Sousa Fragoso³
e Maria Leonor da Silva Carvalho⁴

Resumo: Neste artigo, pretende-se estimar os coeficientes de afetação dos custos às atividades agrícolas, a partir de informação incompleta. Um modelo de desagregação dos custos de fatores de produção variáveis por atividade foi desenvolvido com base na teoria da máxima entropia e foi aplicado à região alentejo, usando os dados da base RICA de 2008. Os resultados obtidos mostram que a teoria da entropia é uma opção flexível e robusta para realizar a estimação de coeficientes de afetação dos custos das atividades num contexto de informação incompleta. Entre as duas variantes utilizadas, o modelo da entropia cruzada generalizada revelou maior capacidade de previsão, uma vez que os coeficientes estimados são mais próximos dos valores reais.

Palavras-chaves: Máxima entropia, custos variáveis, modelo de desagregação, base RICA.

Abstract: *This paper aims to estimate the cost allocation coefficients of agricultural activities based on incomplete information. A model to disaggregate the variable input costs based on the maximum entropy theory was developed. It was applied to the Alentejo region using data from the FADN network of 2008. Results show that entropy theory is a flexible and robust option for estimating cost allocation coefficients of agricultural activities in a context of incomplete information. Between model variants used, the model of generalized cross entropy had the best forecast capacity, once the estimated coefficients are closer to the real values.*

Key-words: *maximum entropy, variable costs, FADN database, disaggregation model.*

Classificação JEL: C6, Q12.

-
1. Trata-se de um artigo cujos autores são de Portugal. Dessa forma, o artigo está escrito em português corrente utilizado em Portugal, em que não há correspondência de termos com o português utilizado no Brasil, mesmo para textos científicos em economia.
 2. Técnica do Gpp, MAM, Lisboa, Portugal. E-mail: socorro@gpp.pt
 3. Departamento de Gestão, CEFAGE, Universidade de Évora. Évora, Portugal. E-mail: rfragoso@uevora.pt
 4. Professora, Departamento de Economia, ICAAM, Universidade de Évora. Évora, Portugal. E-mail: leonor@uevora.pt

1. Introdução

Na União Europeia há uma grande diversidade de estruturas e sistemas de produção e, apesar da crescente tendência para a especialização, a maioria das explorações agrícolas são multiproducto. Para facilitar a análise e para haver comparabilidade das características estruturais e dos resultados económicos das explorações agrícolas dos diferentes Estados Membros, foi desenvolvida uma Rede de Informação de Contabilidades Agrícolas (RICA) com o objetivo de recolher anualmente toda a informação relativa à contabilidade da exploração agrícola, e a Tipologia para a Classificação Comunitária das Explorações Agrícolas (Tipologia) pertencentes à RICA.

Em Portugal, desde os anos 80 que a RICA tem desenvolvido um trabalho de recolha de informação a nível da exploração agrícola, numa amostra constante conforme o Regulamento nº 79/65/CEE que institui a RICA e o Regulamento nº 2143/81/CEE que a melhora. A informação disponível baseia-se nos dados de cada exploração agrícola numa visão contabilística, em que os custos totais por tipo de fator são agregados, não sendo detalhados por atividade.

A competitividade é um objetivo amplamente divulgado e solicitado às empresas do setor agrícola, mas não pode ser alcançada apenas por via dos preços de comercialização de produto, pois as empresas compradoras querem cada vez mais

pagar menos, independentemente da origem do produto. A forma de aumentar a competitividade nesse contexto consiste em melhorar a eficiência (gestão) na utilização dos fatores de produção nas diversas atividades.

A análise económica das atividades agrícolas em resposta às mudanças de medidas de políticas e de mercado, sistemas de produção, tecnologias e competitividade também se baseia na estrutura de custos de produção das atividades. Estas estruturas são inerentes à existência de uma empresa agrícola e esses custos podem ser variáveis, fixos e atribuídos, sendo os primeiros os custos específicos de uma atividade. A informação relativa aos encargos fixos e variáveis de produção por unidade de atividade é extremamente importante, não só ao nível da gestão da empresa agrícola mas também na análise de políticas. O agricultor, apesar de dispor da informação necessária, não consegue detalhá-la na forma contabilística mais amigável para a análise económica da empresa e a gestão das suas atividades, e que permita saber quais as atividades mais interessantes a desenvolver e as que podem ser alteradas. As explorações agrícolas que fazem parte da RICA têm essa informação agregada para a exploração por tipo de fatores de produção e nunca separados por atividade, ou seja, as atividades têm a sua produção individualizada, mas os custos dos fatores de produção estão organizados de uma forma agregada para a totalidade da exploração. Desta forma não

é possível conhecer a estrutura de custos e, por conseguinte, a rentabilidade por atividade.

O uso de técnicas baseadas em métodos econométricos ou na metodologia da máxima entropia pode ser uma solução válida para estimar os encargos com um custo relativamente reduzido. Para tentar responder ao problema da falta de informação relativa aos custos dos fatores por atividade agrícola, este artigo tem como principal objetivo estimar os coeficientes de afetação desses custos às atividades agrícolas, a partir da base de dados da RICA. Para o efeito é desenvolvido um modelo de desagregação dos custos de fatores de produção variáveis por atividade com base na teoria da máxima entropia. Este modelo é aplicado no contexto da região agrária do Alentejo, segundo os dados da RICA de 2008.

Este trabalho encontra-se organizado em seis partes. Na secção 2 faz-se uma breve revisão dos principais métodos utilizados na desagregação de custos, dando especial atenção aos métodos baseados na teoria da máxima entropia. Na secção 3 descreve-se a metodologia e na secção 4 os dados empíricos utilizados. Na secção 5 analisam-se os resultados e finalmente na secção 6 sintetizam-se as principais conclusões.

2. Desagregação de custos de produção e estimação de coeficientes de afetação

A falta de informação sobre os sistemas de produção é um problema determinante que pode ser atenuado com o recurso a métodos de desagregação de dados, nomeadamente dos custos da produção agrícola por atividades. Deste modo o problema da falta de informação transforma-se num problema de desagregação de custos, formalizado para uma amostra de T explorações, que realizam K atividades agrícolas utilizando fatores de produção que dão origem a I itens de custo:

$$X_i^t = \sum_{k=1}^K \alpha_{ik} Y_k^t + u_i^t \quad i = 1, 2, \dots, I \quad \text{e} \quad t = 1, 2, \dots, T \quad (1)$$

X_i^t = Custo do *input* i por exploração t;

α_{ik} = Coeficiente de afetação de custos por *input* i e por produto k;

Y_k^t = Receita total ou produto bruto por produto k e por exploração t;

u_i^t = Resíduo aleatório da distribuição que é específico de cada *input* i e cada exploração t.

Os coeficientes de afetação α_{ik} são os parâmetros desconhecidos da equação (1) e são definidos em termos dos seus valores por actividade agrícola ou produto e por item de custo ou *input* para a globalidade das empresas da amostra objecto de estudo.

Existem vários métodos para estimar os custos específicos das atividades das explorações agrícolas. Estas estimativas são baseadas nas relações que se estabelecem entre os coeficientes de afetação da produção e os respetivos fatores, assumindo a hipótese de que os agentes económicos têm um comportamento de maximização do lucro e tendo em conta as devidas desagregações dos fatores, estimativas de custos de produção e coeficientes de produtividade (JUST *et al.*, 1983; LENCE e MILLER, 1998a, 1998b).

Para se proceder à desagregação dos custos por atividade agrícola através da estimativa de parâmetros, pode recorrer-se às técnicas de regressão linear. No entanto, a fiabilidade dos parâmetros estimados depende do número de empresas ou de explorações agrícolas da amostra. Segundo Love (1999) e Zhang e Fan (2001), quando a informação provém de uma amostra pequena, as técnicas de regressão linear não devem ser utilizadas, pelo que deverão ser adotados outros procedimentos metodológicos.

Por outro lado, Peeters e Surry (2002, 2005) referem que a imposição de restrições contabilísticas, como a que garante a igualdade do rendimento bruto total aos custos totais, dificultam ou inviabilizam mesmo a utilização de algumas técnicas de análise para estimar os coeficientes de desagregação dos custos. Estes autores também se referem às técnicas de regressão linear e à técnica de estimação bayesiana, como as metodologias clássicas mais usadas, apesar dos problemas que ambas comportam na sua aplicação.

Nas técnicas de regressão linear, os principais problemas prendem-se com a possibilidade de obtenção de estimativas negativas e com o facto de todas as equações serem interdependentes (BEWLEY, 1986). O problema da negatividade dos coeficientes pode ser resolvido com a utilização de técnicas de estimação bayesiana ou através do método dos mínimos quadrados com restrições (MOXEY e TIFFIN, 1994). No entanto, estes métodos, para além de não permitirem incorporar as restrições de saldo contabilístico, a sua aplicação empírica é mais complexa.

Golan *et al.* (1994) usaram a máxima entropia (ME) e a mínima entropia cruzada (EC) para derivar uma matriz de contabilidade social. Mais tarde, Golan *et al.* (1997) e Golan *et al.* (2001) utilizaram o método da máxima entropia generalizada (MEG) para estimar os parâmetros de um modelo de regressão. Paris e Howitt (1998) recorreram a uma formulação de MEG para estimar os parâmetros das funções custo das explorações a serem incorporadas na função objetivo de um modelo de programação matemática positiva.

Leon *et al.* (1999) e Peeters e Surry (2002) mostraram que a qualidade de precisão dos coeficientes extraídos de amostras de explorações agrícolas com dados incompletos através dos métodos de entropia é elevada. Estes autores, para além de concluírem que as características estatísticas demonstradas pelos estimadores da entropia são boas, verificaram também que os resultados não apresentam diferenças significativas, quando comparados com uma amostra de encargos observados desagregados por atividade.

Fraser (2000) utilizou a MEG para estimar a procura de carne no Reino Unido e Zhang e Fan (2001) para estimar as tecnologias de produção na agricultura chinesa e para proceder à desagregação dos fatores de produção por província. Garvey e Britz (2002) estimaram a desagregação de custos a partir da contabilidade da RICA também com a MEG e Campbell e Hill (2005, 2006) mostraram como colocar restrições de informação utilizando diferentes vetores de suporte. Howitt e Msangi (2006) recorreram à MEG para estimar funções de produção numa amostra

pequena de 27 explorações no Rio Bravo, México. Rezek e Campbell (2006) para estimar os custos de diversos efluentes/poluentes a partir de dados de painel.

Hansen e Surry (2006) derivaram a procura de quantidades de fatores para diferentes produções na Alemanha utilizando o método de máxima entropia aplicado aos dados das Contas Económicas Regionais. Lips em 2009, partindo da informação da RICA, utilizou a ME para desagregar os custos das várias atividades pelas diversas áreas de culturas arvenses na Suíça. Xavier *et al.* (2010) utilizaram a ME para criar um modelo de desagregação dos dados da utilização da terra, dos efetivos pecuários e de produções. Papalia (2010) também usou a MEG para a estimação de um modelo de interação espacial a partir de dados de painel.

Fragoso *et al.* (2008) utilizaram um modelo EC para proceder à desagregação das superfícies agroflorestais de acordo com zonas agroecológicas pré-definidas para o ecossistema montado no Alentejo.

3. Metodologia

Para proceder à desagregação dos custos de produção da exploração agrícola por atividade a partir da base de dados da RICA utilizou-se o método da ME, método que permite estimar parâmetros sem a imposição de restrições à distribuição da probabilidade e dos erros a partir de amostras pequenas (GOLAN *et al.*, 1996; CAMPBELL e HILL, 2005, 2006). Este método tem também a vantagem de não requerer informação sobre o comportamento do agricultor e de poder integrar no modelo informação adicional que não consta da amostra.

Considerou-se um modelo de MEG, cujos resultados foram melhorados com ajuda de uma distribuição *a priori* numa formulação de entropia cruzada generalizada (ECG). A aplicação desses modelos de ME permite estimar os coeficientes de afetação dos custos dos fatores de produção da exploração agrícola pelas suas atividades atra-

vés de um sistema de equações lineares. Este conjunto de equações não é mais do que as restrições de informação que condicionam as estimativas dos coeficientes de afetação. Os custos dos fatores são assumidos como uma função linear do valor da produção das atividades.

3.1. A Máxima Entropia Generalizada

O modelo MEG tem sido muito utilizado para a resolução de problemas relacionados com a desagregação de fatores. Peeters e Surry (2002, 2005) adotaram este método para estimar os coeficientes de afetação dos custos dos fatores de produção às atividades agrícolas a partir de uma amostra de explorações; Golan *et al.* (2001) e Peeters e Surry (2002, 2005) utilizaram-no tendo introduzido uma formulação Tobit nos seus modelo de ME. Este procedimento permite separar as situações em que a informação está disponível daquelas em que não existe. A sua aplicação é muito útil no caso da aplicação dos modelos de ME ao contexto das explorações agrícolas, na medida em que nem todas as explorações realizam todas as atividades agrícolas disponíveis, correspondendo, por conseguinte, nestes casos, a uma situação de ausência de informação, que é necessário corrigir. Deste modo, tendo em conta a formulação Tobit, as equações de informação na sua forma matricial e vetorial passam a ser as seguintes:

$$X = YA + u = \begin{bmatrix} x_1 > 0 \\ x_2 = 0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} A + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad (2)$$

em que x_1 , y_1 e u_1 dizem respeito aos respetivos valores não nulos da informação observada e x_2 , y_2 e u_2 representam a ausência de informação.

De acordo com Golan *et al.* (1996b) e Peeters e Surry (2002), o modelo estatístico linear pode ser alterado, reparametrizando A e u em função dos seus parâmetros aleatórios p e w_1 e w_2 e dos vetores de suporte z_1 , z_2 , v_1 e v_2 :

$$\begin{bmatrix} x_1 > 0 \\ x_2 = 0 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} z'P + \begin{bmatrix} v'1 & w_1 \\ v'2 & w_2 \end{bmatrix} \quad (3)$$

A seguir apresenta-se a formulação Tobit simplificada do modelo de MEG que foi utilizado para estimar os coeficientes de afetação dos custos dos fatores de produção agrícola.

$$\begin{aligned} \text{Max}(p, w_1, w_2) = & \\ & - \sum_{i=1}^I \sum_{k=1}^K \sum_{m=1}^M p_{ki}^m \ln(p_{ki}^m) - \\ & \sum_{t=1}^T \sum_{i=1}^I \sum_{n=1}^N w_{it}^n \ln(w_{it}^n) - \sum_{t=1}^T \sum_{i=1}^I \sum_{n=1}^N w_{it}^n \ln(w_{it}^n) \end{aligned} \quad (4)$$

Sujeito a:

$$\begin{aligned} X1_i^t = & \sum_{k=1}^K \sum_{m=1}^M p_{ki}^m \cdot z_{ki}^m \cdot y1_k^t + \\ & + \sum_{n=1}^N w_{it}^n \cdot v1_{i,t}^n \quad \forall t1 \text{ e } \forall i \in \{1, \dots, T\} \end{aligned} \quad (5)$$

$$\begin{aligned} X2_i^t = & \sum_{k=1}^K \sum_{m=1}^M p_{ki}^m \cdot z_{ki}^m \cdot y2_k^t + \\ & + \sum_{n=1}^N w_{it}^n \cdot v2_{i,t}^n \quad \forall t2 \text{ e } \forall i \in \{1, \dots, T\} \end{aligned} \quad (6)$$

$$\sum_{m=1}^M p_{k,1}^m = 1 \quad \forall k \in \{1, 2, \dots, I\} \quad (7)$$

$$\begin{aligned} \sum_{n=1}^N w_{it}^n = 1, \\ \forall i \in \{1, 2, \dots, I\} \text{ e } \forall t \in \{1, 2, \dots, T_1\} \end{aligned} \quad (8)$$

$$\begin{aligned} \sum_{n=1}^N w_{it}^n = 1, \\ \forall i \in \{1, 2, \dots, I\}, \forall t \in \{1, 2, \dots, T_2\} \text{ e } T_1 + T_2 = T \end{aligned} \quad (9)$$

$$\sum_{i=1}^I \sum_{m=1}^M p_{ki}^m \cdot z_{ki}^m = 1, \quad \forall k \in \{1, 2, \dots, K\} \quad (10)$$

A equação (4) diz respeito à maximização da entropia conjunta da distribuição de probabilidades dos coeficientes de afetação dos custos dos fatores de produção (p_{ki}^m) e das distribuições de probabilidades dos erros (w_{it}^n e w_{it}^n). O objetivo é maximizar a medida de entropia de Shannon de modo a determinar as probabilidades (p_{ki}^m , w_{it}^n e w_{it}^n) ótimas e únicas que verificam as restrições de coerência da informação (5) e (6), tendo em conta as restrições (7) a (9) relativas às propriedades das distribuições das probabilidades (Jaynes, 1957a; 1957b).

As equações (5) e (6) representam as restrições de consistência da informação, referindo-se a primeira às t_1 explorações com valores positivos em $y1_k^t$ e a segunda às t_2 explorações com valores nulos em $y2_k^t$. Essas equações compreendem a necessidade de consistência de informação com um mínimo de conteúdo, tratando as relações da

equação linear do problema (1) como um sistema independente de equações, em que todos os fatores de produção i são contabilizados em conjunto e em simultâneo. No entanto, os valores positivos são separados dos valores nulos.

As equações (7) a (9) são restrições relativas às distribuições da probabilidade e criam no modelo a normalização dos valores de probabilidade de p_{ki}^m e w_{it}^n relativamente às suas dimensões M e N , respetivamente. Garante-se deste modo que a soma das probabilidades estimadas é igual à unidade.

A equação (10) corresponde a uma condição adicional que estabelece que o somatório dos coeficientes de afetação dos custos dos fatores de produção estimados para cada atividade k seja igual a 1.

3.2. O modelo de mínima entropia cruzada

Para tentar melhorar os resultados das estimativas dos coeficientes de afetação dos custos, utilizou-se uma distribuição a priori exterior à amostra de explorações agrícolas objeto de estudo, recorrendo à formulação ECG. Neste caso, a distribuição a priori utilizada diz respeito à estrutura dos custos utilizada pelo Gabinete de Planeamento e Políticas (GPP) para realizar o apuramento das margens brutas padrão das atividades agrícolas para o ano de 2004. Harris (2002) utilizou técnicas de mínima entropia cruzada para reconciliar dados referentes a diferentes fontes na estimativa da matriz contabilística social regional no México, onde as regiões rurais são diferenciadas por sistemas de produção agrícola com diversas tecnologias. Fragoso *et al.* (2008) mostraram que a utilização da mínima entropia cruzada pode ser um processo muito útil para a desagregação espacial de dados. Xavier *et al.* (2010) utilizaram a mínima entropia cruzada para criar um modelo de desagregação dos dados da agricultura da região Algarve para um nível local numa zona de intervenção florestal.

A abordagem da entropia permite-nos incorporar no modelo de estimação informação adicional de forma explícita, através do princípio da entropia cruzada introduzido por Good (1963).

Esta variante da entropia é muito útil para melhorar a qualidade das estimativas, uma vez que o objetivo é minimizar a discrepância entre as probabilidades estimadas e as distribuições a priori das probabilidades.

Neste caso, o modelo de ECG tem a mesma estrutura de equações que o modelo de MEG, ou seja, inclui as restrições de coerência de informação que permitem relacionar as variáveis dependentes com as variáveis independentes do problema e as restrições das distribuições das probabilidades. A única alteração que se verifica é ao nível da função objetivo que passa a ser a minimização da entropia cruzada entre a distribuição dos coeficientes a estimar e a distribuição dos coeficientes que constituem a distribuição a priori, como se representa a seguir:

$$\begin{aligned} \text{Min}(\alpha, w_1, w_2) = & \sum_{i=1}^I \sum_{k=1}^K \alpha_k^i \ln \left(\frac{\alpha_k^i}{q_k^i} \right) + \\ & \sum_{t=1}^T \sum_{i=1}^I \sum_{n=1}^N w_{1,t,i}^i \ln(w_{1,t,i}^i) + \\ & \sum_{t=1}^T \sum_{i=1}^I \sum_{n=1}^N w_{2,t,i,n}^i \ln(w_{2,t,i,n}^i) \end{aligned} \quad (11)$$

em que α é o coeficiente de afetação dos custos que se pretende estimar e q é o coeficiente conhecido da distribuição a priori utilizada.

3.3. Limites para os intervalos dos vectores de suporte

A escolha dos vectores de suporte, nomeadamente, dos limites do intervalo a que pertencem os seus valores e número de pontos suporte desses vectores é sempre uma tarefa difícil e delicada, sendo necessário ter em conta não só a amplitude do intervalo, como também a simetria da distribuição e por conseguinte da forma como se distribuem os seus valores no espaço.

Golan *et al.* (1996a) consideram que se aumentar a largura do intervalo dos limites do vetor suporte e o número de pontos suporte, pode-se aumentar a informação produzida dos dados estimados. No entanto, Paris e Howitt (1998) defendem que, apesar dos valores dos vetores suporte

z poderem modificar as estimativas de uma forma significativa, não há uma regra para proceder à escolha de um vetor certo para a distribuição.

No que se refere ao vetor de suporte do erro, é muito comum o uso da regra dos três sigmas (3σ), que consiste em considerar o erro centrado na origem e admitir uma amplitude de 3 desvios padrão à direita e de 3 desvios padrão à esquerda (Pukelsheim, 1994). Exemplos da utilização deste procedimento para tratar o erro da distribuição são os trabalhos de Golan *et al.* (1996a), Peeters e Surry (2002), Chakir (2009), Martins *et al.* (2011), Pires *et al.* (2010) e Fragoso e Carvalho (2011).

Peeters e Surry (2002 e 2005) utilizaram, respetivamente, num estudo feito para o Canadá e num estudo feito para França, um vetor suporte z definido no intervalo [0,1], que nesses casos eram os únicos valores teoricamente aceitáveis, onze pontos de suporte ($M=11$) e a regra dos três sigmas na definição do erro com um vetor de suporte v com três pontos de apoio ($N=3$). Fragoso e Carvalho (2011) definiram os limites do vetor suporte z no intervalo [0,1] e consideraram apenas 3 pontos de apoio ($M=3$). Para o erro consideraram a regra dos três sigmas na amplitude do intervalo do vetor v. Fragoso *et al.* (2008) e Van Delden *et al.* (2006) também consideraram o mesmo procedimento, que está de acordo com (GOLAN *et al.*, 1996). O processo de definição do erro seguido por esses autores também foi verificado por Chakir (2009) e Martins *et al.* (2011).

Neste estudo vai-se considerar para o vetor de suporte das probabilidades estimadas $M=3$ pontos e $Z = \{0, 0,5, 1\}$ porque os resultados não pareceram ser sensíveis a outras alternativas de z. No que diz respeito à definição do erro, adotou-se a regra dos três sigmas e estabeleceram-se três pontos de apoio ($N=3$) para o vetor v, i.e., $v = \{-\sigma, 0, \sigma\}$.

3.4. Indicadores de qualidade e precisão

Para avaliar a qualidade e a precisão das estimativas dos modelos de entropia apresentados na secção anterior, considerou-se o cálculo dos indicadores da entropia normalizada, do pseudo- R^2 , dos ganhos de informação obtidos no

processo de desagregação de custos e do desvio absoluto médio. Esses indicadores têm sido utilizados por vários autores para aferir as propriedades estatísticas dos estimadores da entropia e para concluir da sua validade prática, por comparação com dados observados (PEETERS e SURRY, 2002; HOWITT e MSANGI, 2006; FRAGOSO e CARVALHO, 2011).

O indicador de entropia normalizada (S) é definido por Golan *et al.* (1996) como a proporção de incerteza que permanece na estimativa. Este indicador é muito útil para avaliar sobre o desempenho relativo das estimativas dos coeficientes da entropia no que diz respeito à geração de nova informação e à participação da informação inicial contida nos dados nessas estimativas.

Neste caso, com base em Golan *et al.* (1996b) e em Peeters e Surry (2003), considerou-se o cálculo da entropia normalizada dos coeficientes de afetação dos custos dos fatores de produção ($S(\hat{p})$) e o cálculo da entropia normalizada para a estimativa dos erros ($S(\hat{w})$):

$$S(\hat{p}) = \frac{-\hat{p} \ln \hat{p}}{IK \ln(M)}$$

$$S(\hat{w}) = \frac{-\hat{w} \ln \hat{w}}{IK \ln(N)}$$

sendo I o número de fatores de produção ou itens de custo, K o número de atividades agrícolas, T o número de explorações e M e N o número de pontos de apoio considerados para os vetores suporte z e v.

O valor do indicador da entropia normalizada varia entre 0 e 1. O valor zero indica que não há incerteza na estimativa e que a informação dos dados está totalmente refletida nas estimativas, não se tendo por conseguinte produzido qualquer tipo de informação nova no processo de desagregação dos dados. Pelo contrário, se o valor indicado é igual à unidade, a incerteza é máxima e a participação dos dados na informação contida nas estimativas é mínima, indicando que a produção de informação nova é máxima.

Para avaliar a capacidade de estimação do modelo utilizou-se como indicador o Pseudo R^2 ,

que foi também utilizado com o mesmo fim por Peeters e Surry (2002), Fragoso e Carvalho (2011), Fragoso *et al.* (2008), Chakir (2009), Martins *et al.* (2011) e Xavier *et al.* (2010).

O Pseudo R^2 é definido como o quadrado do coeficiente da correlação entre os valores estimados e os valores observados:

$$R_i^2 \equiv \frac{|\sum_{t=1}^T x_t^i \hat{x}_t^i|^2}{|\sum_{t=1}^T \hat{x}_t^i x_t^i|^2}, \quad \text{com} \quad \hat{x}_t^i = \sum_{k=1}^K \hat{\alpha}_k^i y_{k,t}$$

Os valores do indicador Pseudo R^2 variam entre 0 e 1, significando no primeiro caso uma reduzida capacidade de estimação do modelo e no último a capacidade estimação é máxima.

Uma das medidas da qualidade da aderência das estimativas com a realidade e simultaneamente da quantidade de informação produzida no processo de desagregação é o indicador de ganhos de informação desagregada (GID). Este indicador foi estabelecido por Howitt e Reynaud (2003) sob a designação original de Disaggregation Informational Gain (DIG) e entre nós foi utilizado por Xavier *et al.* (2010) e por Fragoso *et al.* (2008). O cálculo deste indicador é dado por:

$$GID = 1 - \frac{\hat{EC}}{EC} = 1 - \frac{\sum_{k=1}^K \sum_{i=1}^I \hat{\alpha}_k^i \ln\left(\frac{\hat{\alpha}_k^i}{\alpha_k^i}\right)}{\sum_{k=1}^K \sum_{i=1}^I \alpha_k^i \ln\left(\frac{\alpha_k^i}{\alpha_k^i}\right)}$$

em que \hat{EC} é a entropia cruzada estimada e EC é a entropia cruzada observada. A \hat{EC} mede a entropia entre a distribuição estimada e a distribuição agregada. Quando o valor da \hat{EC} é zero, a distância é nula e não existe heterogeneidade nos dados. A EC mede a entropia cruzada dos coeficientes da informação observada no nível desagregado e no nível agregado, dando a heterogeneidade de informação observada quando desagregada.

Segundo Howitt e Reynaud (2003), o GID deverá aumentar à medida que se ganha informação com a desagregação. O ganho é nulo quando a desagregação é uniforme e sem heterogeneidade. Quando o GID é 1, atinge o seu valor máximo, e indica que toda a informação foi recuperada.

No que diz respeito à validação e avaliação da utilidade prática do modelo, optou-se por realizar a comparação das estimativas com os dados observados, através do cálculo do desvio absoluto médio (DAM) por fator de produção ou item de custo (i) e por atividade agrícola (k):

$$DAM_k^i = \left| \frac{\alpha_k^i - \hat{\alpha}_k^i}{\alpha_k^i} \right| \times 100$$

com α_k^i representando os dados observados que são comparados com os resultados estimados.

4. Implementação empírica do modelo

Para tentar responder ao problema da falta de informação relativa aos custos dos fatores por atividade agrícola, desenvolveu-se um modelo de desagregação dos custos de fatores de produção variáveis por atividade com base na teoria da máxima entropia. Este modelo é aplicado no contexto da região agrária do Alentejo, segundo os dados da RICA de 2008.

Uma análise preliminar dos dados ajuda a organizá-los de acordo com as categorias de custo específico e de rendimento, selecionar explorações, atividades e custos a estudar. Do ficheiro de resultados da RICA consta um extenso número de atividades, para cada uma das regiões de Portugal, tendo-se selecionado o Alentejo como região de estudo, pela dimensão física elevada das explorações agrícolas e por existir também alguma especialização das atividades produzidas.

Tabela 1. Nº de explorações por classes de OTE e MBT no Alentejo

OTE	MBPT (□)	%	nº explorações	%
Culturas Arvenses	2312814	22	40	16
Horticultura Intensiva	305837	3	10	4
Culturas Permanentes	1077999	10	36	15
Herbívoros	4519264	43	115	47
Granívoros	279879	3	2	1
Policultura	372920	4	7	3
Polipecuária	533736	5	9	4
Mistas Culturas e Pecuária	1145318	11	28	11
Total	10547768	100	247	100

Fonte: RICA, 2010

A Tabela 1 apresenta a distribuição de explorações por classes de orientação técnico económica (OTE) e margem bruta padrão total (MBPT), na amostra RICA de 2008 para a região Alentejo. Da sua análise, verifica-se que as OTE que agregam mais explorações agrícolas são a OTE dos Herbívoros (47%), a OTE das Culturas Arvenses (16%) e a OTE das Culturas Permanentes (15%). Estas três OTE representam 78% do total de explorações da amostra para o Alentejo e contribuem com 75% para a MBPT da região. Cada uma dessas OTE apresenta um número de explorações que permite tratar cada uma delas como uma amostra autónoma, o que permite obter resultados por OTE.

Foi extraída uma amostra de conveniência de explorações agrícolas da base RICA 2008, por OTE. No sentido de tornar essa amostra mais equilibrada em termos de culturas, áreas e produções, desenvolveram-se alguns critérios, que passam também por apostar principalmente nas culturas com alguma área produtiva (culturas que sejam comercialmente ativas) e nas explorações com alguma dimensão (para o Eurostat as empresas têm de ter mais de 5 ha de superfície agrícola utilizada). Também se analisaram e eliminaram as explorações que apresentavam problemas de ordem técnica tais como as explorações que se encontravam a fazer implantação de culturas permanentes (em período de carência de rendimentos), as empresas que apresentavam coeficientes económicos negativos, como as mar-

gens brutas, as explorações que tinham tido problemas de sanidade animal e consequente vazio sanitário e as empresas com uma especificidade determinante no sistema de produção que as tornavam diferentes, como a produção de cavalos (equídeos), com grandes e cuidadosos regimes alimentares e os suínos e as aves com uma vertente industrial.

De acordo com a metodologia usada para calcular as margens brutas padrão, os itens de custo considerados para as atividades vegetais foram sementes e plantas, fertilizantes, fitofármacos e outros encargos com as culturas. Para as atividades animais esses itens foram rações compradas, alimentos forrageiros comprados e produzidos na exploração, assistência veterinária e outros encargos com a pecuária. O valor da margem bruta é definido em euros por atividade (vegetal ou animal). A unidade usada para as culturas é o hectare e para os animais é cabeça normal (CN).

Para validar o modelo de desagregação de custos por atividade é necessário dispor dos coeficientes dos custos observados e desagregados por fator de produção e produto. Para esse efeito utilizou-se a estrutura de custos de referência das atividades utilizada no cálculo das margens brutas padrão (MBP) de 2004.

Nas culturas forrageiras e pastagens, a área utilizada é a que consta da ficha de exploração, uma vez que esta área não consta do ficheiro de resultados da RICA. Também as produções destas atividades, sendo auto utilizadas, não apre-

sentam valor na base de dados, pelo que o valor da sua produção teve de ser estimado através do Valor de Produção Padrão (VPP) de 2008.

As 29 empresas selecionadas na OTE 1- Culturas Arvenses têm uma área média de 97 ha de superfície agrícola, com um mínimo de 31 ha e um máximo de 309 ha de área total, mostrando uma grande variação na sua dimensão física. As atividades que apresentam valores de receita mais elevados são as culturas horto-industriais, milho grão e milho forragem. As produções médias alcançadas pelas empresas (€ de receita/ha produzidos) apresentam um intervalo relativamente grande em quase todas as atividades. As atividades que mais contribuem para a receita global da OTE são as horto-industriais (44%). Quando se analisam os encargos por hectare também se verifica que os gastos são muito diferentes, sendo os intervalos com maior amplitude, os apresentados pelos fitofármacos e pelos fertilizantes. A margem bruta também apresenta alguma variabilidade mas bem menor que os encargos.

As empresas selecionadas na OTE 3- Culturas Permanentes apresentam uma área média de 33 ha, com mínimo de 6 ha e um máximo de 149 ha. As atividades que apresentam maior receita média por ha são a fruticultura seguida de viticultura. O peso da contribuição de cada atividade no rendimento da OTE 3 é de 48% por parte da viticultura e 40% da olivicultura. No que se refere à área, a olivicultura é dominante (67%), seguida da vinha (20%) e o trigo com 8%. De notar que a fruticultura ocupa apenas 2% da área, no entanto apresenta os melhores resultados médios por ha (6076€), seguida da viticultura (3447€). Quando se observam os encargos, verifica-se que as variações são grandes nos outros encargos e fitofármacos. A margem bruta apresenta a menor das variações mínimas.

As empresas orientadas para herbívoros com atividade agrícola (OTE 4) apresentam áreas diversas desde 60 ha a 413 ha, sendo a média de 163 ha. Ao nível das atividades, as maiores receitas médias por ha são provenientes de milho forrageiro, trigo e pastagens temporárias.

No entanto, a maior contribuição em termos de receita para a OTE é dada pela atividade de prados temporários com 33% do valor, seguido de outras forragens e milho forrageiro com 21% e prados permanentes com 16%, num total de 91%, mostrando claramente a prioridade da empresa em fornecer alimentação aos seus animais.

A maior área média está concentrada nas pastagens permanentes que condicionam cerca de 52% da área total envolvida e as outras culturas forrageiras apresentam um peso de 21%.

As empresas orientadas para herbívoros e com atividade vegetal também apresentam flutuações bastante elevadas nos encargos como sucede, com os outros encargos e os fitofármacos.

As empresas orientadas para herbívoros (animal) apresentam quantidades de cabeças normais diversas desde 6 CN a 269 CN, com uma média de 72 CN. Ao nível das atividades, a maior receita média por CN é proveniente da atividade de bovinos de leite, sendo os bovinos de carne a atividade que mais contribui para as receitas da OTE, com cerca de 46%. São estes os que apresentam o maior efetivo médio (43CN), seguido dos ovinos com 25CN. Por outro lado, essas mesmas espécies contribuem para com 64% e 34% do peso do efetivo da OTE. Os encargos por CN apresentam variações muito grandes.

5. Resultados

Neste ponto apresentam-se os resultados dos modelos de Tobit-MEG e Tobit-ECG aplicados a quatro amostras de explorações agrícolas da RICA no Alentejo para o ano de 2008, relativas às OTE das culturas arvenses, das culturas permanentes, dos herbívoros com componente vegetal e dos herbívoros componente animal.

O modelo Tobit-ECG foi utilizado com o objetivo de melhorar as estimativas do modelo Tobit-MEG, fazendo uso de priores de informação que não estão contidos nas amostras objeto de estudo. Neste caso utilizou-se como distribuição *a priori*, $q_{i,k}$, a estrutura de custos utilizada pelo GPP do

Ministério da Agricultura em Portugal para calcular as margens brutas padrão do ano 2004 por atividade agrícola (Rosário, 2012).

Os resultados dos modelos Tobit-MEG e Tobit-ECG são apresentados para cada uma das amostras de explorações agrícolas consideradas e são analisados, primeiro em termos dos indicadores de entropia normalizada e de capacidade de prescrição (pseudo-R²), depois em termos dos ganhos de informação obtidos no processo de desagregação e por último em termos da sua validade prática através da comparação entre os resultados estimados e os dados observados para os coeficientes de afetação dos custos dos fatores por atividade agrícola, através do cálculo do DAM.

Na Tabela 2 apresentam-se os resultados dos modelos MEG e ECG relativos aos indicadores da entropia normalizada e do pseudo-R² das explorações agrícolas por OTE.

A entropia normalizada $S(\hat{p})$ apresenta os valores mais elevados para o MEG para os casos das OTE das culturas arvenses (0,708) e dos herbívoros-animal (0,755). Nestes casos, os valores obtidos de entropia normalizada são relativamente próximos da unidade, o que denota uma percentagem razoável de criação de nova informação no processo de desagregação de custos.

No modelo de ECG, como seria de esperar, o valor da entropia normalizada $S(\hat{p})$ é menor,

i.e., mais perto de zero, registando-se também os valores mais elevados nas OTE das culturas arvenses (0,307) e dos herbívoros-animal (0,403). Tanto no modelo MEG como no modelo ECG, a percentagem de criação de informação no processo de desagregação regista os valores mais baixos nos casos das culturas permanentes (0,450 e 0,253) e dos herbívoros-vegetal (0,489 e 0,198).

No que diz respeito à componente do erro, os valores obtidos para entropia normalizada $S(\hat{w})$ são em todos os casos e para ambos os modelos muito próximos da unidade (entre 0,980 e 0,993).

Como foi referido anteriormente, o pseudo-R² foi utilizado para avaliar a capacidade de estimação do modelo, i.e., a proporção da variância que é explicada pelo modelo. Neste caso, o pseudo-R² foi calculado para cada tipo de custo separadamente. A margem bruta é o item que apresenta os valores do pseudo-R² mais próximos da unidade, nomeadamente, nas OTE das culturas arvenses (0,916 e 0,968), das culturas permanentes (0,995 e 0,997) e dos herbívoros-vegetal (0,987 e 0,983). É também de referir os bons resultados que se obtiveram neste indicador para o caso da OTE dos herbívoros-animal nos itens de custo relativos aos concentrados e às forragens. Os itens de custos que apresentam os valores de pseudo-R² inferiores são os outros custos em especial nas OTE das culturas permanentes (0,293 e 0,337) e dos herbívoros-vegetal (0,231 e 0,239). Como seria de

Tabela 2. Entropia normalizada e pseudo-R² para os modelos de MEG e ECG por OTE

	OTE		OTE		OTE		OTE	
	Culturas arvenses		Culturas. Permanentes		Herbívoros – vegetal		Herbívoros – animal	
	MEG	ECG	MEG	ECG	MEG	ECG	MEG	ECG
Entropia Normalizada	0,708	0,307	0,450	0,253	0,489	0,198	0,755	0,403
Entropia Normalizada do erro	0,981	0,991	0,974	0,981	0,975	0,980	0,992	0,993
Pseudo R²								
Sementes e Plantas	0,867	0,928	0,636	0,643	0,501	0,479	-	-
Fertilizantes	0,822	0,864	0,845	0,934	0,590	0,582	-	-
Pesticidas	0,753	0,803	0,893	0,920	0,229	0,357	-	-
Outros Custos	0,773	0,800	0,293	0,337	0,231	0,239	0,654	0,654
Margem Bruta	0,916	0,968	0,995	0,997	0,987	0,983	0,895	0,920
Concentrados	-	-	-	-	-	-	0,940	0,931
Forragens	-	-	-	-	-	-	0,801	0,784

Fonte: Resultados do modelo.

Tabela 3. Ganhos de informação desagregada (GID) dos modelos de MEG e ECG por OTE

	OTE		OTE		OTE		OTE	
	Culturas arvenses		Culturas. Permanentes		Herbívoros – vegetal		Herbívoros – animal	
	MEG	ECG	MEG	ECG	MEG	ECG	MEG	ECG
GID total	0,87	0,99	0,97	0,99	0,92	0,99	0,81	0,99
GID parcial								
Trigo	0,87	1,00	0,98	0,98	0,97	0,99	-	-
Milho	0,99	0,99	-	-	-	-	-	-
Arroz	0,98	0,99	-	-	-	-	-	-
O. cereais	0,98	1,00	0,88	0,90	0,97	0,98	-	-
Proteaginosas	0,82	1,00	-	-	-	-	-	-
Horto industriais	0,99	0,99	-	-	-	-	-	-
Fruticultura	-	-	0,98	0,98	-	-	-	-
Oleaginosas	0,87	1,00	-	-	-	-	-	-
Olivicultura	0,83	1,00	0,99	1,00	0,90	1,00	-	-
Viticultura	-	-	0,99	0,99	-	-	-	-
Pastagens permanentes	0,84	1,00	-	-	0,95	0,98	-	-
Pastagens pobres	-	-	-	-	0,82	1,00	-	-
Pastagens temporárias	0,90	1,00	-	-	1,00	0,99	-	-
Milho forrageiro	0,91	1,00	-	-	0,99	0,99	-	-
Outras forragens	0,92	1,00	-	-	0,99	1,00	-	-
Bovinos de carne	-	-	-	-	-	-	0,96	0,99
Bovinos de leite	-	-	-	-	-	-	0,96	0,99
Ovinos	-	-	-	-	-	-	0,75	1,00
Caprinos	-	-	-	-	-	-	0,78	1,00

Fonte: Resultados do modelo.

esperar, os valores obtidos para o pseudo- R^2 são sempre melhores no modelo ECG do que modelo MEG, confirmando a superioridade deste modelo em termos da capacidade de estimação.

Os ganhos de informação desagregada (GID) total para o modelo MEG variam entre 0,81 no caso dos herbívoros componente animal e 0,97 no caso das culturas permanentes. Nas restantes duas OTE, os valores do GID total são 0,87 nas culturas arvenses e 0,92 nos herbívoros componente vegetal (Tabela 3). Estes resultados, em termos dos ganhos totais de informação no processo de desagregação, podem ser considerados bons, sendo os resultados obtidos com o modelo ECG superiores, uma vez que atingem em todas as OTE objeto de estudo 0,99.

O GID parcial evidencia, na generalidade das atividades envolvidas, os ganhos que ocorrem parcialmente e como influenciam o GID.

No caso da OTE das culturas arvenses para o modelo de MEG, o GID parcial varia entre 0,82 nas proteaginosas e 0,99 no milho e nas horto-

-industriais, sendo visíveis dois grupos atividades, um com valores de GID menos elevados composto pelas proteaginosas (0,82), a olivicultura (0,83), a pastagem permanente (0,84), trigo (0,87) e oleaginosas (0,87), e outro que inclui todas as atividades com GID superior a 0,9.

No modelo de ECG, os ganhos de informação são superiores aos obtidos para o modelo de MEG e são quase totais, uma vez que o valor do indicador é na maior parte dos casos igual à unidade.

Para as culturas permanentes, os ganhos de informação total são elevados, sendo o valor deste indicador 0,977 no modelo de MEG e 0,994 no modelo de ECG. Quando se analisa os GID parciais, apenas os outros cereais registam um GID abaixo dos 0,9 e por conseguinte um comportamento algo diferente das restantes atividades.

Nos ganhos de informação desagregada total para os herbívoros na componente vegetal é 0,924 no modelo de MEG, sendo este valor superior no modelo de ECG (0,992). No modelo de MEG, as pastagens pobres e a olivicultura regis-

tam os GID menores (0,82 e 0,90). As restantes atividades apresentam valores de GID de 1,0 ou nas redondezas de 0,98 e 0,99. Aqui também os resultados obtidos com o modelo de ECG se mostraram superiores.

A OTE dos herbívoros componente animal apresenta o pior desempenho em termos de ganho da informação desagregada total. No caso do modelo de MEG o GID total é 0,81. Com o modelo de ECG, os valores do GID são muito superiores quando comparados com os do modelo MEG. Nos ganhos de informação parcial, os piores valores relativos são obtidos nos ovinos (0,75) e nos caprinos (0,78). No modelo ECG, o GID parcial assume os valores de 0,99 ou de 1.

Na Tabela 4 apresentam-se os resultados das estimativas dos coeficientes de afetação dos custos dos fatores de produção às atividades agrícolas e o respetivo desvio absoluto médio (DAM) obtidos com o modelo de MEG.

Para as culturas arvenses, verifica-se uma coerência generalizada dos valores estimados. Dos 60 coeficientes de afetação α estimados, mais de um terço apresenta um DAM que varia entre zero e 5% e cerca de metade regista um valor inferior a 20%. Os itens de custo relativos às sementes e plantas nos casos do arroz, proteaginosas e outras forragens apresentam valores de desvios elevados. Os fertilizantes apresentam também valores bastante elevados do DAM nos casos do arroz, horto-industriais, olivicultura e oleaginosas. Os pesticidas têm desvios elevados nos outros cereais e olivicultura. Os outros custos têm desvios consideráveis nas diversas atividades, devido a ser um grupo de custos muito diversificado. As margens brutas apresentam desvios entre 17% nas horto-industriais e 77% nas proteaginosas.

No caso das explorações da OTE Culturas Permanentes os coeficientes de afetação obtidos

com o modelo de MEG apresentam na generalidade dos casos valores coerentes com os valores reais. Dos 25 coeficientes de afetação estimados praticamente um terço apresenta um DAM entre zero e 1,1%. Apenas os pesticidas nos outros cereais apresentam um valor de desvio muito elevado. Por outro lado, as margens brutas também apresentam desvios elevados nos casos do trigo e outros cereais. A análise descritiva do DAM no modelo MEG revela que os desvios máximos são de um modo geral inferiores aos que se obtiveram na OTE Culturas Arvenses, com exceção dos pesticidas que também apresentam desvios e coeficientes de variação elevados.

No caso da OTE dos herbívoros componente vegetal, também se obtiveram com o modelo de MEG estimativas dos coeficientes de afetação dos custos dos fatores relativamente próximos dos valores respetivos valores reais. Aqui também dos 40 coeficientes de afetação estimados mais de um terço tem um DAM igual ou abaixo dos 5%. No entanto, verificam-se desvios significativos nas estimativas dos coeficientes dos fertilizantes em diversas atividades, assim como nos outros custos nas atividades de outros cereais, pastagens permanentes, milho forragem e outras forragens e nos pesticidas nos outros cereais, na olivicultura e nas outras forragens.

Os coeficientes de afetação obtidos com o modelo de MEG para a OTE Herbívoros na sua componente animal apresentam desvios significativos em relação aos valores reais. Dos 16 parâmetros estimados apenas 12,5% têm um DAM inferior a 5%. As piores estimativas dos coeficientes de afetação verificam-se nos bovinos, que apresentam valores de DAM significativamente elevados em todos os itens de custos. Os caprinos registam também uma DAM significativamente elevado no item de custo relativo às forragens.

Tabela 4. Coeficientes de afetação (α) e DAM obtidos com o modelo de MEG por OTE

	Semente/Plantas		Fertilizantes		Pesticidas		Outros Custos		Margem Bruta	
	α	DAM	α	DAM	α	DAM	α	DAM	α	DAM
OTE – Culturas arvenses:										
Trigo	0,177	60,9	0,206	17,6	0,154	18,9	0,185	0,0	0,279	36,6
Milho	0,162	32,5	0,236	24,2	0,139	39,0	0,184	67,3	0,278	20,6
Arroz	0,190	216,7	0,210	133,3	0,237	5,2	0,131	0,8	0,232	51,7
Outros cereais	0,153	17,7	0,225	2,2	0,143	615,0	0,180	100,0	0,299	44,6
Proteaginosas	0,198	120,0	0,200	0,0	0,195	0,0	0,201	0,0	0,206	77,4
Horto industriais	0,078	13,3	0,152	204,0	0,086	4,4	0,064	220,0	0,621	17,2
Oleaginosas	0,164	2,5	0,231	670,0	0,175	0,0	0,154	413,3	0,275	64,7
Olivicultura	0,196	0,0	0,204	308,0	0,196	553,3	0,193	0,0	0,212	76,7
Past.permanente	0,200	0,0	0,200	81,8	0,190	0,0	0,198	120,0	0,213	73,4
Past.Temporária	0,127	0,0	0,199	5,2	0,119	0,0	0,154	3,8	0,401	36,3
Milho forrag.	0,124	55,0	0,211	75,8	0,130	0,0	0,135	237,5	0,399	47,5
Out. forragens	0,196	292,0	0,202	34,7	0,195	0,0	0,197	0,0	0,210	72,7
OTE – Culturas Permanentes:										
Trigo	0,143	58,7	0,301	1,1	0,241	19,6	0,017	0,0	0,297	154,7
Out. cereais	0,062	74,6	0,232	45,0	0,181	1370,1	0,047	90,6	0,478	478,4
Fruticultura	0,001	0,0	0,118	15,8	0,050	0,0	0,001	99,6	0,831	97,1
Olivicultura	0,003	0,0	0,114	32,5	0,070	0,0	0,007	93,7	0,806	28,0
Viticultura	0,010	0,0	0,037	77,7	0,083	0,0	0,016	59,3	0,854	58,6
OTE – Herbívoros componente vegetal:										
Trigo	0,103	58,5	0,333	311,1	0,095	32,6	0,029	0,0	0,439	38,5
O. cereais	0,103	37,3	0,307	338,6	0,062	342,9	0,022	266,7	0,505	39,5
Olivicultura	0,142	0,0	0,293	2563,6	0,151	843,8	0,076	0,0	0,339	65,2
Past. permanente	0,070	0,0	0,222	753,8	0,046	0,0	0,012	140,0	0,651	32,8
Pastagens pobres	0,203	0,0	0,209	0,0	0,200	0,0	0,176	0,0	0,211	78,9
Past. temporária	0,002	0,0	0,047	24,2	0,003	0,0	0,001	94,1	0,947	2,2
Milho forrag.	0,022	42,1	0,070	141,4	0,018	0,0	0,009	350,0	0,880	5,5
Out. forragens	0,055	161,9	0,144	300,0	0,014	250,0	0,017	1600,0	0,771	17,7
OTE – Herbívoros componente animal:										
Bovinos de carne	0,332	149,6	0,062	264,7	0,204	264,3	0,402	49,4		
Bovinos de leite	0,411	152,1	0,056	460,0	0,138	160,4	0,395	49,0		
Ovinos	0,328	0,0	0,054	68,8	0,214	22,3	0,403	49,2		
Caprinos	0,274	0,0	0,176	309,3	0,281	72,4	0,270	66,0		

Fonte: Resultados do modelo.

Os resultados das estimativas dos coeficientes de afetação dos custos dos fatores de produção às atividades agrícolas e o respectivo desvio absoluto médio (DAM) obtidos com o modelo de ECG são apresentados na Tabela 5. Neste caso os valores do DAM revelam que em geral os coeficientes de afetação estimados são mais próximos dos valores reais do que as estimativas obtidas com o modelo de MEG.

Da observação dos desvios no modelo de ECG para as culturas arvenses percebe-se um

melhor relacionamento dos coeficientes de afetação, sendo os desvios em relação aos valores reais menores. Dos 60 coeficientes estimados praticamente metade apresentam um DAM inferior a 20% e 22% têm um DAM entre 0 e 5%. Os desvios mais elevados verificam-se nos itens dos fertilizantes e dos outros custos nas horto-industriais (137,4% e 165%), no item dos pesticidas no milho (203%) e nos outros cereais (180%) nas outras forragens (119,8%), devido a estas atividades

Tabela 5. Coeficientes de afetação (α) e DAM obtidos com o modelo de ECG por OTE

	Semente/Plantas		Fertilizantes		Pesticidas		Outros Custos		Margem Bruta	
	α	DAM	α	DAM	α	DAM	α	DAM	α	DAM
OTE – Culturas arvenses:										
Trigo	0,113	28,0	0,23	4,4	0,163	37,0	0,000	0,0	0,494	3,8
Milho	0,166	42,5	0,255	72,1	0,149	203,0	0,180	63,6	0,249	17,9
Arroz	0,082	11,9	0,116	58,3	0,267	93,1	0,074	43,1	0,461	2,6
Outros cereais	0,100	36,4	0,238	32,1	0,029	180,1	0,140	55,6	0,494	2,8
Proteaginosas	0,091	28,2	0,000	0,0	0,000	0,0	0,000	0,0	0,909	4,1
Horto industrias	0,082	31,8	0,106	137,4	0,083	58,1	0,053	165,0	0,676	7,9
Oleaginosas	0,125	39,9	0,038	66,7	0,000	0,0	0,033	10,0	0,803	13,2
Olivicultura	0,000	0,0	0,051	6,3	0,031	61,6	0,000	0,0	0,918	0,6
Past.permanente	0,000	0,0	0,104	19,8	0,000	0,0	0,084	6,7	0,811	11,8
Past.Temporária	0,000	0,0	0,121	24,6	0,000	0,0	0,099	38,1	0,780	45,5
Milho forrag.	0,046	56,8	0,089	9,9	0,000	0,0	0,029	27,5	0,836	17,2
Out. forragens	0,049	22,0	0,149	15,3	0,01	119,8	0,020	0,0	0,772	2,8
OTE – Culturas Permanentes:										
Trigo	0,128	18,5	0,065	73,0	0,311	161,4	0,000	0,0	0,496	4,2
Out. cereais	0,040	74,6	0,131	27,3	0,126	1117,2	0,026	1,0	0,677	40,9
Fruticultura	0,000	0,0	0,101	85,2	0,000	0,0	0,000	2,0	0,898	6,0
Olivicultura	0,001	0,0	0,121	152,3	0,054	0,0	0,007	3,0	0,817	11,5
Viticultura	0,008	0,0	0,028	46,8	0,064	0,0	0,016	4,0	0,884	0,1
OTE – Herbívoros componente vegetal:										
Trigo	0,068	56,7	0,265	10,2	0,058	51,3	0,004	47,3	0,605	27,1
O. cereais	0,082	47,9	0,226	25,5	0,009	13,1	0,014	91,9	0,669	39,3
Olivicultura	0,000	0,0	0,050	4,3	0,015	21,8	0,006	37,5	0,930	0,7
Past. permanente	0,028	0,0	0,146	68,1	0,019	0,0	0,010	94,7	0,796	9,7
Pastagens pobres	0,000	0,0	0,000	0,0	0,000	0,0	0,000	0,0	1,000	0,0
Past. temporária	0,000	0,0	0,022	86,3	0,000	0,0	0,000	99,9	0,978	82,4
Milho forrag.	0,022	79,4	0,051	48,4	0,017	0,0	0,009	88,9	0,901	26,3
Out. forragens	0,049	22,0	0,088	31,9	0,001	87,3	0,017	67,8	0,845	12,6
OTE – Herbívoros componente animal:										
Bovinos de carne	0,254	16,0	0,046	55,9	0,124	220,9	0,577	9,6		
Bovinos de leite	0,403	48,9	0,052	17,7	0,116	211,5	0,429	31,8		
Ovinos	0,000	0,0	0,034	83,0	0,163	27,1	0,802	19,4		
Caprinos	0,000	0,0	0,146	42,5	0,156	39,3	0,698	10,1		

Fonte: Resultados do modelo.

corresponderem a agregados de culturas, com comportamentos diferentes em produção.

No modelo de ECG das culturas permanentes os coeficientes de afetação apresentam valores consideráveis de DAM para os fertilizantes na olivicultura (152,3%), na fruticultura (85,2%) e no trigo (73%) e de novo para os pesticidas nos outros cereais (1117,2%), que apresentam um valor muito elevado, e no trigo (161,4%). Neste caso, dos 25 coeficientes estimados 32% têm um

DAM entre zero e 5% e quase metade está abaixo dos 20%.

A análise do DAM dos coeficientes de afetação estimados com o modelo de ECG nos herbívoros na sua componente vegetal mostra em geral que os valores são aceitáveis. Cerca de 35% dos coeficientes estimados têm um DAM entre 0 e 5%. Os desvios mais significativos verificam-se nos itens de custo relativos aos fertilizantes na pastagem temporária (86,3%), aos pesticidas nas

outras forragens (87,3%), aos outros custos na pastagem temporária (99,9%), na pastagem permanente (94,7%), nos outros cereais (91,9%) e no milho forrageiro (88,9 %).

No caso do modelo de ECG relativo aos herbívoros componente animal os valores encontrados para os coeficientes de afetação são também mais normalizados e mais próximos dos valores reais. Apenas os outros custos continuam a registrar valores de DAM elevados no caso dos bovinos de carne (220,9%) e dos bovinos de leite (211,5%).

De modo geral, os valores obtidos com a intervenção de indicadores de qualidade, precisão e estimação têm comportamentos muito diferentes, apesar de, em alguns casos, serem muito semelhantes aos que são apresentados em diversos trabalhos como Fragoso e Carvalho (2011), Peeters e Surry (2002), nomeadamente, para as entropias normalizadas e para o pseudo- R^2 . Também os valores obtidos para os GID são relativamente semelhantes aos apresentados por Fragoso e Carvalho (2011), Xavier *et al.* (2010) e Howitt e Reynaud (2003). Os valores referentes aos coeficientes de afetação e aos DAM também são relativamente semelhantes aos apresentados por Fragoso e Carvalho (2011).

As principais dificuldades sentidas foram, tal como referem Peeters e Surry (2002), problemas com as categorias heterogêneas como as outras oleaginosas e outros cereais onde se encontram grupos de culturas. Por outro lado, as variações ou melhor a heterogeneidade entre as empresas e, nomeadamente, a variabilidade entre os custos de fatores também levaram a que existisse alguma disparidade entre os valores estimados e os valores reais observados, como aconteceu com Fragoso e Carvalho (2011). Não obstante, os mesmos autores avançam com alguns valores para analisar o DAM na calibração de modelos, considerando que o modelo se encontra bem validado para valores inferiores a 15%. No entanto, um DAM superior a 15% pode ser considerado razoável, mas mostra que o modelo precisa de ser melhorado. Howitt e Msangi (2006) consideram aceitáveis valores de DAM entre 1,5% e 40,9%.

Ao analisar-se detalhadamente as características das explorações verifica-se que há alguns

pontos que eventualmente poderão ser melhorados. Na Orientação Técnico-Económica Culturas Arvenses, deveriam estar apenas culturas arvenses e não horto-industriais, culturas permanentes e culturas forrageiras anuais e permanentes, que são muito diferentes em área, em tecnologia/sistema de produção e rendimento e por conseguinte também em termos dos custos de produção. As áreas das explorações estão inseridas num intervalo muito grande, entre 31 e 309 ha. As receitas deste grupo têm uma amplitude muito grande, sendo as mais elevadas nas explorações de horto-industriais e as mais baixas nas culturas e prados permanentes. O produto bruto mostra as diferenças acentuadas que existem na mesma atividade, sobressaindo o grau de intensificação da agricultura e conseqüentemente o uso maior ou menor de fatores de produção como sementes e plantas, fertilizantes, pesticidas e outros encargos, com reflexo direto na margem bruta. Nas outras duas classes de Orientação Técnico-Económica consideradas para o estudo, permanecem também as mesmas dificuldades que foram referidas para as Culturas Arvenses.

6. Conclusão

A falta de informação desagregada relativamente a fatores de produção é um problema generalizado na área das ciências agrárias. A utilização de inquéritos diretos e pessoais junto do agricultor, dispendiosos e demorados, é a fonte de informação. O uso de técnicas e métodos alternativos para estimar os coeficientes de afetação dos custos às atividades agrícolas, a partir de informação incompleta, pode ser uma experiência válida para estimar os custos variáveis de produção unitários da exploração a um custo significativamente mais baixo.

A informação sobre as explorações constantes das bases de dados da RICA respeita ao total dos custos por tipo de fatores, não sendo desagregada por atividade. A obtenção dos custos variáveis unitários de produção por unidade de atividade é extremamente importante quer para

a gestão da exploração quer na perspetiva da análise da política agrícola.

O desenvolvimento de um modelo baseado na teoria da máxima entropia permitindo obter, a partir da base de dados RICA, os encargos de produção desagregados por atividade, mostrou-se muito útil. Este método apresenta a grande vantagem de ultrapassar dificuldades relacionadas com a informação incompleta, e ainda a sua flexibilidade e facilidade de manuseamento e também a existência de valores nulos na amostra.

Os resultados apresentados mostram que se conseguiu desagregar a informação a partir dos modelos de máxima entropia generalizada e da entropia cruzada generalizada com a estimação de coeficientes de afetação por atividade. A aplicação de critérios de qualidade de precisão e de previsão mostram que os modelos aplicados dão boas indicações no sentido da desagregação dos fatores e que se encontram num limiar muito perto de outros autores que se basearam nas mesmas questões do sector agrícola, com os mesmos métodos e modelos e com a utilização dos mesmos critérios. A comparação dos resultados dos modelos de máxima entropia generalizada e mínima entropia cruzada mostra que a utilização da entropia cruzada conduz a uma melhoria substancial dos resultados.

No entanto, há que salientar que nem todos os resultados podem ser considerados bons, nomeadamente com a análise feita com o DAM, há alguns valores menos interessantes que precisam ser melhorados. A melhoria do modelo pode requerer mais algum trabalho para solucionar os desvios percebidos: amostras mais alargadas, amostras com dimensão física semelhante, não utilização de critérios de OTE, não regionalização, com cruzamento de outra informação como painéis de opinião, líderes de opinião e valores provenientes de consulta bibliográfica.

Apesar de algumas limitações nos resultados, a metodologia da entropia apresentou-se como um instrumento de trabalho flexível e robusto para ajudar a compensar falhas de informação, como é o caso dos custos dos fatores de produção desagregados por atividade agrícola.

7. Referências bibliográficas

CAMPBELL, R. e CARTER HILL, R. A Monte Carlo study of the effect of design characteristics on the inequality restricted maximum entropy estimator, *Review of Applied Economics*, v. 1, p. 53-84, 2005.

_____. e CARTER HILL, R. Imposing parameters inequality restrictions using the principle of maximum entropy", *Journal of Statistical Computation and Simulation*, v. 76, p. 985-1000, 2006.

CHAKIR, R. Spatial downscaling of agricultural land use data: an economic approach using cross entropy, *Land Economics*, v. 85, n. 2, p. 238-251, 2009.

FRAGOSO, R., MARTINS, M.B. e LUCAS, M.R. Disaggregated soil allocation data using a Minimum Cross Entropy model, *WSEAS Transactions on Environment and Development*, v. 9, n. 4, p. 756-766, 2008.

_____. e CARVALHO, M.L. Estimation of Farm Cost Allocation Coefficients at Farm Level Using an Entropy Approach, *CEFAGE-UE Working-Papers*, 2011/21.

FRASER, I. An application of maximum entropy estimation the demand for meat in the United Kingdom, *Applied Economics*, v. 32, n. 4, p. 5-59, 2000.

GARVEY, E. e BRITZ, W. Estimation of Input Allocation from EU Farm Accounting Data using Generalized Maximum Entropy, *CAPRI Working Paper 02-01*, U.Ireland & Bonn, 2002.

GOLAN, A., JUDGE, G. e ROBINSON, S. Recovering information from incomplete or partial multisectorial economic data, *Review of Economics and Statistics*, v. 76, p. 541-551, 1994.

_____. JUDGE, G. e MILLER D. *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. New York: John Wiley Editions, 1996a.

_____. JUDGE, G. e PERLOFF M. A Maximum Entropy Approach to Recovering Information From Multinomial Response Data. *Journal of the American Statistical Association*, v. 91, n. 434, p. 841-853, 1996b. Disponível em: <<http://www.jstor.org/stable/2291679>>.

_____. KARP, S. E PERLOFF, M. Estimation and inference with censored and ordered multinomial response data, *Journal of Econometrics*, v. 73, p. 23-52, 1997.

_____. PERLOFF, M. e SHEN, Z. Estimating a demand system with the non-negativity constraints: Mexican meat demand, *Review of Economics and Statistics*, LXXXIII, p. 541-551, 2001.

- HANSEN, H. e SURRY, Y. Estimating the cost allocation for German agriculture: an application of the maximum entropy methodology, Conference paper, *46th Annual Conference of German Association of Agricultural Economists*, October 4-6, 2006.
- HARRIS, R. Estimation of a Regionalized Mexican Social Accounting Matrix: using Entropy Techniques to Reconcile Disparate Data Sources, *TMD discussion paper n°97 WDC. USA*, 2002.
- HAZELL, P e NORTON, R. *Mathematical Programming for Economic Analysis in Agriculture*, NY: Mac Millan P.C. USA, 1986.
- HOWITT, R. e REYNAUD, A. Spatial Disaggregation of Agricultural Production Data using Maximum Entropy, *European Review of Agriculture Economics*, v. 30, n. 3, p. 359-387, 2003.
- _____. e MSANGI, S. Estimating Disaggregate Production Functions: An Application to Northern Mexico, Draft paper, 2006.
- JAYNES, E.T. Information theory and statistic mechanics, *Physics Review*, v. 106, p. 620-630, 1957a.
- _____. Information theory and statistic mechanics, *Physics Review*, v. 108, p. 171-190, 1957b.
- JUST, R., ZILBERMAN, D. e HOCHMAN, E. Estimation of Multicrop Production Functions, *American Journal of Agricultural Economics*, v. 65, p. 770-780, 1983.
- LENCE, H.L. e MILLER, D. Estimation of Multi-Output Production Functions with Incomplete Data: A Generalized Cross Entropy Approach, *European Review of Agricultural Economics*, v. 25, p. 188-209, 1998a.
- _____. e MILLER, D. "Recovering Output-Specific Inputs from Aggregated Input Data: A Generalized Cross Entropy Approach, *American Journal of Agricultural Economics*, v. 80, p. 852-867, 1998b.
- LEON, Y., PEETERS, L., QUINQUE, M. e SURRY, Y. The use of maximum entropy to estimate input-output coefficients from regional accounting data, *Journal of Agricultural Economics*, v. 50, p. 425-439, 1999.
- LIPS, M. Full product costs on base of farm accountancy data by means of maximum entropy, Paper presented to the International Association of Agricultural Economists Conference, Beijing, China. August, 2009.
- LOVE, H. A. Conflicts Between Theory and Practice in Production Economics, *Amer. J. Agr.Econ.*, v. 81, p. 696-702, 1999.
- MARTINS, M.B., FRAGOSO, R. e XAVIER, A. Spatial disaggregation of Agricultural Data: A maximum Entropy Approach, *JP Journal of Biostatistics*, v. 5, n. 1, p. 1-16, 2011.
- MOXEY, A. e TIFFIN, R. Estimating linear production coefficients from farm business survey data: A note, *Journal of Agricultural Economics*, v. 45, p. 381-385, 1994.
- PAPALIA, R. Generalized Maximum entropy Estimation of Spatial panel data interaction models, *WSEAS Transactions on Business and Economics*, 14, v. 9, p. 322-331, 2010.
- PARIS, Q. e HOWITT, R.E. An analysis of ill-posed production problems using maximum entropy, *American Journal of Agricultural Economics*, n. 80, p. 124-138, 1998.
- PEETERS, L. e SURRY, Y. Farm cost allocation based on the maximum entropy methodology- The case of Saskatchewan crops farms-Agriculture and Agri-Food Canada Strategic Policy Branch. Publication 2121/E, Project 02-011-tp, Ottawa, Ontario, 2002.
- _____. e SURRY, Y. Estimation d'un modèle à paramètres variables par la méthode d'entropie croisée généralisée et application à la répartition des couts de production en agriculture, In: *Actes des Journées de Méthodologie Statistique*, 2005.
- PUKELSHEIM, F. The Three Sigmas Rule, *American Statistician*, v. 48, p. 88-91, 1994.
- REZEK, J. e CAMPBELL, R. Cost estimates for multiple pollutants: A maximum entropy approach, *Energy Economics*, v. 29, p. 503-519, 2006. Disponível em: <<http://www.sciencedirect.com>>.
- ROSÁRIO, M.S. *Um modelo de desagregação de encargos de produção variáveis da base RICA por actividade*, Dissertação de mestrado, Universidade de Évora, Évora, 2012.
- VAN DELDEN, H. e LUJA, P. Integration of multi-scale dynamic spatial models for land use change analysis and assessment of land degradation and socio-economic processes, In: Conference on Soil protection strategy - needs and approaches for policy support, Poland, 2006.
- XAVIER, A., MARTINS, M.B. e FRAGOSO, R. Combined disaggregation of agricultural land uses, livestock numbers and crops' production: an entropy approach, *Advances in Mathematical and Computacional methods*, p. 192-198, 2010.
- ZHANG, X. e FAN, S. Estimation Crop-Specific Production Technologies in Chinese Agriculture: a Generalized Maximum Entropy Approach, *American Journal of Agricultural Economics*, v. 83, n. 2, p. 378-388, 2001. Disponível em: <<http://www.jstor.org/stable/1244680>>.