Cadavez, Vasco A. P.; Henningsen, Arne

Working Paper

The Use of Seemingly Unrelated Regression (SUR) to Predict the Carcass Composition of Lambs

FOI Working Paper, No. 2011/12

Provided in Cooperation with:
Department of Food and Resource Economics (IFRO), University of Copenhagen

INSTITUTE OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

FOI Working Paper

The Use of Seemingly Unrelated Regression (SUR) to Predict the Carcass Composition of Lambs

Vasco A.P. Cadavez
Arne Henningsen

2011 / 12

FOI Working Paper 2011 / 12

The use of seemingly unrelated regression (SUR) to predict
the carcass composition of lambs

Authors: Vasco A.P. Cadavez, Arne Henningsen

Institute of Food and Resource Economics

University of Copenhagen

Rolighedsvej 25

DK 1958 Frederiksberg DENMARK

www.foi.life.ku.dk

The Use of Seemingly Unrelated Regression (SUR) to Predict the Carcass Composition of Lambs

Vasco A. P. Cadavez
Mountain Research Center (CIMO)
Instituto Politécnico de Bragança

Arne Henningsen
Institute of Food and Resource Economics
University of Copenhagen

Abstract

The aim of this study was to develop and evaluate models for predicting the carcass composition of lambs. Forty male lambs of two different breeds were included in our analysis. The lambs were slaughtered and their hot carcass weight was obtained. After cooling for 24 hours, the subcutaneous fat thickness was measured between the 12th and 13th rib and the total breast bone tissue thickness was taken in the middle of the second sternebrae. The left side of all carcasses was dissected into five components and the proportions of lean meat, subcutaneous fat, intermuscular fat, kidney and knob channel fat, and bone plus remainder were otained. Our models for carcass composition were fitted using the SUR estimator which is novel in this area. The results were compared to OLS estimates and evaluated by several statistical measures. As the models are intended to predict carcass composition, we particularly focussed on the *PRESS* statistic, because it assesses the precision of the model in predicting carcass composition. Our results showed that the SUR estimator performed better in predicting LMP and IFP than the OLS estimator. Although objective carcass classification systems could be improved by using the SUR estimator, it has never been used before for predicting carcass composition.

Keywords: Carcass, Quality, Ordinary least squares, Seemingly unrelated regression.

1. Introduction

Generally speaking, good quality carcasses should present a reduced amount of fat, but there must be enough fat to guarantee a good presentation of the carcasses and for protecting the meat during the refrigeration period. Carcasses with excess fat are undesirable, because this leads to higher production costs and compels meat traders to remove the fat before selling the meat. However, fat plays an important role in the meat sensory characteristics, and a minimum content of fat is needed in order to maximize the meat palatability (Wood 1995; Ferguson 2004). Thus, a carcass with an optimum composition will fetch the highest price, but whenever the carcass composition moves away from the optimum, its value will suffer.

The success of the meat industry relies on its ability to deliver meat products that satisfy the consumers' requirements (Cortez, Portelinha, Cadavez, Rodrigues, and Teixeira 2006). Thus, an accurate system of carcasses classification is of great importance to the meat industry (Kongsro, Roe, Kvaal, Aastveit, and Egelandsdal 2009; Rius-Vilarrasa, Bünger, Maltin, Matthews, and Roehe 2009), since it is the base for fair payments to producers, as well as the communication of consumers' needs through the supply chain (Rius-Vilarrasa *et al.* 2009). Therefore, researchers have dedicated much effort into developing reliable prediction models of carcasses composition, and several research studies (e.g.

Lambe, Navajas, Schofield, Fisher, Simm, Roehe, and Bunger 2008; Hopkins, Ponnampalam, and Warner 2008; Hopkins 2008; Cadavez 2009) have been conducted to develop an objective system for the classification of carcasses to be applied at the slaughter-line. The results attained by Hopkins (2008) and Cadavez (2009) indicate that the lean meat proportion (LMP) of lamb carcasses can be predicted by simple models using the hot carcass weight (HCW) and fat depth measurements as explanatory variables.

A common feature of published work concerning the prediction of carcass composition is the use of single-equation models. In this approach, several independent equations are estimated separately by ordinary least squares (OLS) and the estimated parameters are used to predict the proportions of muscle, fat, and bone of carcasses. However, this assumption of independence is not supported by biological knowledge, and it is well known that carcass compositional traits are correlated both phenotypically and genotypically. As the proportions of the different carcass tissues are correlated, it is expected that the equations for predicting these will be interrelated. Hence, we can expect that the single-equation approach will be inefficient from a statistical point of view (see e.g. Judge, Hill, Griffiths, Lutkepohl, and Lee 1988).

A set of equations which share a common error structure with non-zero covariance is said to be contemporaneously correlated. Zellner (1962) developed the co-called “Seemingly Unrelated Regression” (SUR) estimator that accounts for these contemporaneous correlations and allows the p dependent variables to have different sets of explanatory variables. The SUR method estimates the parameters of all equations simultaneously, so that the parameters of each single equation also take the information provided by the other equations into account. This results in greater efficiency of the parameter estimates, because additional information is used to describe the system. These efficiency gains increase with increasing correlation among the error terms of the different equations (Judge *et al.* 1988), as well as with larger sample size and higher multi-collinearity between the regressors (Yahya, Adebayo, Jolayemi, Oyejola, and Sanni 2008). In the case of models for predicting carcass composition, the SUR method can be used to estimate all parameters of all equations simultaneously, whilst the correlations among the carcass tissues are taken into account. However, in spite of these elegant proprieties, the SUR method has (to our knowledge) not yet been used for estimating carcass composition prediction models.

The aims of this study were to compare alternative models for simultaneously predicting the lean meat proportion (LMP), subcutaneous fat proportion (SFP), intermuscular fat proportion (IFP), bone plus remainder proportion (BP), and kidney knob and channel fat proportion (KCFP) of lamb carcasses, and to compare the efficiency of the OLS and SUR estimators.

2. Material and methods

2.1. Data

Forty male lambs of Churra Galega Bragançana (n = 22) and Suffolk (n = 18) breeds were randomly selected from the experimental flock of the Escola Superior Agrária de Bragança. The lambs were slaughtered, and their carcasses were weighed approximately 30 min after slaughter in order to obtain the hot carcass weight (HCW). After chilling at 4ºC for 24 hours, the carcasses were halved through the centre of the vertebral column, and the kidney knob and channel fat were removed and weighed. During quartering of the carcasses, the subcutaneous fat thickness (C12, mm) between the 12th and 13th rib was measured with a caliper, and the total breast bone tissue thickness (E2, mm) was taken with a sharpened steel rule in the middle of the 2nd sternebrae according to Delfa, González, and Teixeira (1996).

The left side of all carcasses was dissected into muscle, subcutaneous fat, inter-muscular fat, and bone plus remainder (major blood vessels, ligaments, tendons, and thick connective tissue sheets associated with muscles). The carcasses' lean meat proportion (LMP), subcutaneous fat proportion (SFP), intermuscular fat proportion (IFP), bone plus remainder proportion (BP), and kidney and knob channel fat proportion (KCFP) were calculated as a percentage of the total tissues in the carcasses.

2.2. Statistical analysis

Three multiple equations models were developed to simultanously predict the LMP, SFP, IFP, BP and KCFP, and all statistical analyses were undertaken using the software "R" (R Development Core Team 2011) with the add-on package "systemfit" (Henningsen and Hamann 2007). The fitting quality of each multiple equations model was evaluated by the McElroy coefficient of determination (R_M^2), and the fitting quality of single equations was evaluated by the coefficients of determination of estimation (R_e^2), standard errors of the estimate (SEE), and by the standard errors (SE) of the estimated parameters.

After estimating the "full" models by OLS and SUR, all explanatory variables that had a parameter with a marginal level of significance ("P value") larger than 0.20 were removed.

All models were validated using a leaving-one-out cross-validation procedure (Montgomery 1997). This procedure repeats the statistical analysis N times, where N is the number of observations and at each replication a different observation is excluded from the estimation. Hence, in each replication, $N-1$ observations are used for estimating the model and then the dependent variable of the omitted observation is predicted based on the explanatory variables of the observation and the estimated coefficients. The average precision of these out-of-sample predictions was evaluated by computing the so-called predicted residuals sum of squares (PRESS) statistic and the coefficient of determination of prediction (R_p^2). The normality of the residuals was evaluated using the Shapiro-Wilk test.

Ordinary least squares

The general approach of multivariate single-equation regression models requires that there is only one dependent variable in each regression, i.e.

$$y_i = X_i \beta_i + \varepsilon_i, \tag{1}$$

where y_i is the a vector of the N observations of the ith dependent variable, X_i is an $N \times k_i$ matrix of the regressors of the ith equation (including potentially a column of ones), β_i is the vector of the k_i parameters of the ith equation, k_i is the number of regressors (including potentially a constant) of the ith equation, and ε_i is the vector of error terms of the ith equation, which is assumed to be normally distributed. The OLS estimator assumes that all coefficients in the model are unknown and are estimated from data by $\beta_i^{OLS} = (X_i' X_i)^{-1} X_i' y_i$.

If the parameters of each equation are estimated separately by OLS, a potential correlation between the equations is not taken into account. Hence, it is implicitly assumed that the error terms are not contemporaneously correlated, i.e. $E(\varepsilon_{it}\varepsilon_{jt}) = 0 \; \forall \, i \neq j$, where subscripts i and j indicate the equation and subscript t denotes the observation.

Seemingly unrelated regression

Zellner (1962) developed the Seeming Unrelated Regression (SUR) estimator for estimating models with $p > 1$ dependent variables that allow for different regressor matrices in each equation (e.g. $X_i \neq X_j$) and account for contemporaneous correlation, i.e. $E(\varepsilon_{it}\varepsilon_{jt}) \neq 0$. In order to simplify notation, all equations are stacked into a single equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & X_p \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \tag{2}$$

that can be re-written as $Y = X\beta + \varepsilon$, where the $Y = \left(y'_1, y'_2, \dots, y'_p\right)'$ is a vector of all stacked dependent variables, X is a block diagonal design matrix with the i^{th} design matrix X_i on the ii^{th} block, $\beta = \left(\beta'_1, \beta'_2, \dots, \beta'_p\right)'$ is the vector of the stacked coefficient vectors of all equations, the total number of parameters estimated for all p submodels is $K = \sum_{i=1}^{p} k_i$, and $\varepsilon = \left(\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_p\right)'$ is the vector of the stacked error vectors of all equations.

The same estimates as by separate single-equation OLS estimations can be obtained by an OLS estimation of the entire system of equations, i.e. $\beta^{OLS} = (X'X)^{-1} X'y$. The SUR estimator that accounts for interrelations between the single submodels can be obtained by $\beta^{SUR} = \left[X'\Omega^{-1}X\right]\left[X'\Omega^{-1}Y\right]$, where Ω^{-1} is a weighting matrix based on the covariance matrix of the error terms Σ. This covariance matrix $\Sigma = [\sigma_{ij}]$ has the elements $\sigma_{ij} = E\left[\varepsilon_{in}\varepsilon_{jn}\right]$, where ε_{in} is the error term of the n^{th} observation of the i^{th} equation. Finally, the inverse of the weighting matrix can be calculated by $\Omega = \Sigma \otimes I_N$, where I_N is an $N \times N$ identity matrix and $\otimes$ denotes the Kronecker product. However, as the true error terms ε are unknown, they are often replaced by observed residuals, e.g. obtained from OLS estimates, i.e. $\hat{\varepsilon}_i = y_i - X_i\beta_i^{OLS}$ so that the elements of the covariance matrix can be calculated by[1]

$$\hat{\sigma}_{ij} = \frac{\hat{\varepsilon}'_i \hat{\varepsilon}_j}{N}. \tag{3}$$

Thus, a SUR model is an application of the generalized least squares (GLS) approach and the unknown residual covariance matrix is estimated from the data.

Models for carcass composition

The OLS estimates are obtained while ignoring any correlation between the error terms of different equations. However, if the error terms are contemporaneously correlated, as is most likely in the case of carcass composition studies, the estimation procedure should take this into account. In this case, the SUR estimator leads to efficient parameter estimates (Yahya *et al.* 2008).

[1] Other possibilities for calculating the covariance matrix of the error terms are described in, e.g. Henningsen and Hamann (2007).

Our base model for carcass composition ("C12+E2") consists of five single equations to simultaneously predict the LMP, SFP, IFP, BP and KCFP of lamb carcasses:

$$\begin{aligned} LMP &= \alpha_0 + \alpha_1\, HCW + \alpha_2\, C12 + \alpha_3\, E2 + \varepsilon_1 \quad (4)\\ SFP &= \beta_0 + \beta_1\, HCW + \beta_2\, C12 + \beta_3\, E2 + \varepsilon_2 \quad (5)\\ IFP &= \gamma_0 + \gamma_1\, HCW + \gamma_2\, C12 + \gamma_3\, E2 + \varepsilon_3 \quad (6)\\ BP &= \delta_0 + \delta_1\, HCW + \delta_2\, C12 + \delta_3\, E2 + \varepsilon_4 \quad (7)\\ KCFP &= \theta_0 + \theta_1\, HCW + \theta_2\, C12 + \theta_3\, E2 + \varepsilon_5 \quad (8) \end{aligned}$$

In this model α_0, α_1, α_2, and α_3 are the regression coefficients and ε_1 is the error term in the model for the lean meat proportion (LMP); β_0, β_1, β_2, and β_3 are the regression coefficients and ε_2 is the error term in the model for the subcutaneous fat proportion (SFP); γ_0, γ_1, γ_2, and γ_3 are the regression coefficients and ε_3 is the error term in the model for the intermuscular fat proportion (IFP), δ_0, δ_1, δ_2, and δ_3 are the regression coefficients and ε_4 is the disturbance term in the model for the bone plus remainders proportion (BP), and θ_0, θ_1, θ_2, and θ_3 are the regression coefficients and ε_5 is the disturbance term in the model for the kidney knob and channel fat proportion (KCFP).

Furthermore, our analysis includes the sub-model "C12" with $\alpha_3 = \beta_3 = \gamma_3 = \delta_3 = \theta_3 = 0$, i.e. without the explanatory variable E2, and the sub-model "E2" with $\alpha_2 = \beta_2 = \gamma_2 = \delta_2 = \theta_2 = 0$, i.e. without the explanatory variable C12.

3. Results and Discussion

3.1. Descriptive statistics

The mean, standard deviations, minimum and maximum along with the correlation coefficients among HCW, C12, E2, LMP, SFP, IFP, BP and KCFP are shown in Table 1. HCW varied from 8.0 to 15.0 kg (CV around 16%) and C12 varied from 0.25 to 5.4 mm and presented the highest CV (> 65%). From the carcass tissues, the LMP had the lowest CV (4.7%), followed by the BP with a CV of 7.7%, whilst FP had the highest CV (> 23%). These results are in agreement with the low variability observed in the LMP by Silva (2001) and Cadavez (2009), and this small variation in the carcasses' LMP was pointed out as the main constraint to prediction models with high determination coefficient (Fortin and Sherestha 1986; Silva 2001; Cadavez 2009), since this statistic is highly influenced by the variation observed in the dependent variable (Chatterjee, Hadi, and Price 2000).

The HCW had low and positive correlation ($r = 0.13$) with C12 measurement and medium and positive correlation ($r = 0.52$) with E2. The HCW had low and negative correlation ($r = -0.14$) with LMP, medium and positive correlations with SFP ($r = 0.47$) and with KCFP ($r = 0.31$). The LMP was highly and negatively correlated with SFP ($r = -0.84$) and with KCFP ($r = -0.73$). The SFP had medium and positive correlation with IFP ($r = 0.56$) and high and positive correlation with KCFP ($r = 0.82$). The C12 had medium and negative correlation with LMP ($r = -0.36$), but high and positive correlation with SFP ($r = 0.63$). The E2 had high and negative correlation with LMP ($r = -0.61$), and high and positive correlations with SFP ($r = 0.77$) and with KCFP ($r = 0.75$).

3.2. Results of OLS and SUR estimations

The estimated parameters and summary statistics for the three models "C12", "E2", and "C12+E2" are presented in Table 2. Given that the same regressors were used in each equation, these estimates

Table 1: Descriptive statistics and correlations among HCW, C12, E2, LMP, SFP, IFP, BP, and KCFP

	Descriptive statistics				Correlations						
Variable	Mean	SD	Min	Max	C12	E2	LMP	SFP	IFP	BP	KCFP
HCW, kg	12.2	1.98	8.0	15.0	0.13	0.52	-0.14	0.20	0.47	-0.55	0.31
C12, mm	1.3	0.59	0.35	2.5	1	0.51	-0.36	0.63	0.25	-0.41	0.42
E2, mm	16.4	3.46	7.8	24.3		1	-0.61	0.77	0.52	-0.52	0.75
LMP, %	61.4	2.86	57.7	66.8			1	-0.84	-0.69	0.14	-0.73
SFP, %	4.9	1.73	1.5	8.2				1	0.56	-0.44	0.82
IFP, %	9.1	1.85	3.6	12.9					1	-0.63	0.55
BP, %	23.3	1.78	20.2	27.5						1	-0.48
KCFP, %	1.4	0.50	0.38	2.2							1

(including standard errors) can be obtained either by OLS or by SUR. As the dependent variables sum up to 100% at each observation and exactly the same regressors were used in each equation, the covariance matrix of the residuals becomes singular. Hence, the weighting matrix of the SUR estimator cannot be inverted so that a SUR estimation of all five equations becomes infeasible. However, an arbitrary equation can be dropped and the system can be estimated with the remaining four equations. After the estimation, the parameters of the dropped equation can be retrieved by the "adding-up" restriction, i.e. the intercepts of all five equations have to sum to 100% and the parameters of each explanatory variable (except for the constant) have to sum to zero. The variances and covariances of the parameters of the omitted equation can be calculated by the delta method. The estimated parameters and their variances and covariances do not usually depend on the equation that is dropped. However, as all five equations can be estimated by OLS and these estimates coincide with SUR estimates in our current model specification, there is no advantage of using SUR over OLS for estimating these models.

As expected, the general model "C12+E2" gives the best overall fit, indicated by the highest McElroy-R^2 (0.476), followed by model "E2" (0.399), while model "C12" presents the lowest McElroy-R^2 (0.292). These results indicate that total breast bone tissue thickness (E2) explains a higher proportion of carcass composition than subcutaneous fat thickness (C12), but both measures together explain the largest proportion. This means that both of the carcass fat depth measurements, C12 and E2, are relevant determinants of tissue proportions in lamb carcasses, which is in accordance with the results attained by Cadavez (2009). Our results also confirmed those attained by Hopkins *et al.* (2008) and Hopkins (2008), who showed that HCW alone is unable to explain the LMP.

When looking at the goodness of fit of the individual equations, it becomes apparent that model "C12" (i.e. with explanatory variables HCW and C12) has the lowest coefficients of determination of estimation $\left(R_e^2\right)$ in four out of five equations (LMP, SFP, IFP, and KCFP). For instance, model "C12" can only explain a very limited variation of LMP ($R_e^2 = 0.140$), while model "E2" ($R_e^2 = 0.425$) performs much better, while the general model "C12+E2" ($R_e^2 = 0.425$) is not (noticeably) better than model "E2".

As the correlation between the two fat measurements (C12 and E2) is moderately high ($r = 0.51$), the general model "C12+E2" is somewhat plagued by multicollinearity between its regressors. This results in less precise parameter estimates and can be seen by increased standard errors and lower coefficients of determination of prediction$\left(R_p^2\right)$. In fact, the standard errors of the regression coefficients and the standard errors of the estimates (SEE) are often larger in the general model "C12+E2" than in the specific model "E2". Moreover, the predictive ability of model "E2" measured by the

Table 2: Estimation results of the three basic models

Dependent var.		LMP	LMP	LMP	SFP	SFP	SFP	IFP	IFP	IFP	KCFP	KCFP	KCFP	BP	BP	BP
Model		C12	E2	C12+E2	C12	E2	C12+E2	C12	E2	C12+E2	C12	E2	C12+E2	C12	E2	C12+E2
Interc.	Estim.	65.2	67.0	67.1	1.27	0.348	0.102	3.27	2.69	2.67	0.133	-0.221	-0.227	30.1	30.2	30.4
	SE	2.79	2.31	2.35	1.39	1.08	1.00	1.68	1.62	1.65	0.463	0.350	0.357	1.436	1.50	1.465
	Pr(>|t|)	<2e-16	<2.2e-16	<2.2e-16	0.378	0.750	0.919	0.059	0.105	0.114	0.775	0.553	0.528	<2.2e-16	<2e-16	<2.2e-16
HCW	Estim.	-0.128	0.368	0.361	0.099	-0.250	-0.205	0.412	0.250	0.254	0.065	-0.030	-0.029	-0.449	-0.338	-0.382
	SE	0.222	0.211	0.217	0.111	0.099	0.093	0.134	0.147	0.152	0.037	0.032	0.033	0.114	0.137	0.135
	Pr(>|t|)	0.569	0.090	0.105	0.376	0.016	0.033	0.004	0.098	0.103	0.085	0.355	0.389	0.000	0.018	0.008
E2	Estim.		-0.617	-0.604		0.462	0.376		0.203	0.195		0.119	0.116		-0.166	-0.083
	SE		0.121	0.143		0.056	0.061		0.085	0.100		0.018	0.022		0.078	0.089
	Pr(>|t|)		9.9e-06	0.000		8.0e-10	4.0e-07		0.022	0.059		1.4e-07	4.9e-06		0.041	0.355
C12	Estim.	-1.70		-0.125	1.81		0.829	0.581		0.074	0.325		0.023	-1.02		-0.801
	SE	0.741		0.718	0.371		0.306	0.446		0.503	0.123		0.109	0.382		0.447
	Pr(>|t|)	0.028		0.863	2.1e-05		0.010	0.201		0.884	0.012		0.834	0.011		0.081
R_e^2		0.140	0.425	0.425	0.415	0.658	0.716	0.252	0.323	0.323	0.238	0.576	0.576	0.411	0.373	0.425
R_p^2		0.051	0.375	0.341	0.351	0.622	0.669	0.164	0.239	0.220	0.164	0.544	0.523	0.346	0.303	0.338
SEE		2.72	2.23	2.25	1.36	1.04	0.961	1.64	1.56	1.58	0.452	0.337	0.342	1.401	1.45	1.404

McElroy-R^2 values (based on SUR estimations with an arbitrary equation removed): model "C12" = 0.292, model "E2" = 0.399, model "C12+E2" = 0.476
F-Tests(Theil 1971, p. 314):
- restricting model "C12+E2" to model "C12": F = 8.448 with 5 degrees of freedom, P-value = $3.355 \cdot 10^{-7}$
- restricting model "C12+E2" to model "E2": F = 1.0059 with 5 degrees of freedom, P-value = 0.4157

coefficient of determination of prediction$\left(R_p^2\right)$ is even better than the predictive ability of the general model "C12+E2" for three out of five equations (LMP, IFP, KCFP).

As models "C12" and "E2" are nested in model "C12+E2", we can apply an F-Test on the general model "C12+E2" to test the restrictions implied by models "C12" and "E2". While model "C12" is clearly rejected in favour of the general model "C12+E2", model "E2" is not significantly worse than the general model "C12+E2". Hence, given the costs of taking two measures of carcass fat depth (C12 and E2), model "E2" is probably the most cost-effective prediction model for slaughterhouse applications since it is based on a single fat measurement (E2) and is not significantly worse in predicting lamb carcass composition than the more costly model "C12+E2". As slaughterhouses would only implement a new prediction model if it can be done easily and at low cost, the inclusion of a second fat measurement in the prediction model cannot be justified because of the resulting economic cost.

The correlations between the residuals of our "preferred" model "E2" are shown in Table 3. This table indicates that several equations are highly or moderately interrelated. For instance, the first equation (LMP) is highly interdependent with the second equation (SFP, $r = -0.709$) and the third equation (IFP, $r = -0.670$) and moderately interdependent with the fourth equation (KCFP, $r = -0.505$). As the OLS approach does not take these interrelationships into account, the OLS estimates should be inefficient (not as precise as possible). However, as we use exactly the same explanatory variables in each equation, OLS estimates coincide with SUR estimates and hence, are not negatively affected by the contemporaneous correlation of the error terms.

Table 3: Residuals correlations for model "E2" estimated by ordinary least squares method (OLS)

	LMP	SFP	IFP	KCFP	BP
LMP	1	-0.709	-0.670	-0.505	-0.189
SFP		1	0.445	0.569	-0.242
IFP			1	0.337	-0.446
KCFP				1	-0.229
BP					1

Given that some of the estimated parameters of our "preferred" model "E2" are statistically non-significant and that the inclusion of non-significant explanatory variables generally reduces the precision of the estimates, we tried to improve the precision of our estimates by removing non-significant regressors. Thus, we re-estimated model "E2" after removing all regressors that were not significant at the 20% level. We chose a rather high threshold as we wanted be cautious and avoid removing relevant explanatory variables.[2] As this adjusted version of model "E2", say "E2a", no longer has the same explanatory variables in all equations, OLS estimates differ from SUR estimates and all five equations can be included in the SUR regression. The OLS and SUR estimates as well as summary statistics are shown in Table 4. An F-Test applied to model "E2" shows that the three parameters that have been removed in model "E2a" are jointly not statistically significant. Hence, model "E2a" is not significantly worse than model "E2". The Shapiro-Wilk statistic shows that for both estimators (OLS and SUR) the single-equation residuals have zero mean and follow a normal distribution (data not shown). As the McElroy-R^2 is not intended for OLS regressions and the McElroy-R^2 values for model "E2a" are—in contrast to the models shown in Table 2—based on all five equations, we cannot make reasonable comparisons using the McElroy-R^2 of model "E2a" here. Given that the OLS estimation ignores interrelations between equations, the OLS estimates differ between model "E2" and

[2] In fact, we would remove exactly the same explanatory variables if we had any threshold between 11% and 35%.

Table 4: Estimation results of model "E2a"

Dependent var.		LMP	LMP	SFP	SFP	IFP	IFP	KCFP	KCFP	BP	BP
Estimation method		OLS	SUR	OLS	SUR	OLS	SUR	OLS	SUR	OLS	SUR
Interc.	Estim.	67.0	67.3			2.69	2.60			30.2	30.1
	SE	2.31	1.61			1.62	1.44			1.50	1.45
	Pr(>|t|)	<2.2e-16	<2.2e-16			0.105	0.079			<2e-16	<2e-16
HCW	Estim.	0.368	0.216	-0.229	-0.159	0.250	0.318			-0.338	-0.376
	SE	0.211	0.169	0.073	0.061	0.147	0.136			0.137	0.133
	Pr(>|t|)	0.090	0.208	0.003	0.013	0.098	0.025			0.018	0.007
E2	Estim.	-0.617	-0.523	0.467	0.417	0.203	0.158	0.084	0.084	-0.166	-0.137
	SE	0.121	0.105	0.054	0.046	0.085	0.080	0.003	0.003	0.078	0.076
	Pr(>|t|)	9.9e-06	1.4e-05	1.4e-10	3.8e-11	0.022	0.055	2.2e-16	2.2e-16	0.041	0.082
R_e^2		0.425	0.413	0.657	0.649	0.323	0.317	0.533	0.533	0.373	0.371
R_p^2		0.375	0.382	0.636	0.631	0.239	0.242	0.534	0.534	0.303	0.302
SEE		2.23	2.25	1.03	1.04	1.56	1.56	0.344	0.344	1.445	1.448

McElroy-R^2 values (based on estimations with all 5 equations): OLS = 0.240, SUR = 0.588

F-Test (Theil 1971, p. 314) for restricting model "E2" to this model: F = 0.5935 with 3 degrees of freedom, P-value = 0.62

model "E2a" only if an explanatory variable is removed in the respective equation, i.e. in the equations for SFP and KCFP. In contrast, the SUR estimator accounts for contemporaneous correlations among the equations and hence, the SUR estimates of all equations differ between model "E2" and model "E2a". Furthermore, the efficiency of the SUR estimates compared to the OLS estimates is expected to increase in the presence of highly correlated covariates (Yahya *et al.* 2008), which is a common feature of carcass composition data. Since all tissues are taken in the same experimental unit (the carcass), the measurements are correlated with each other (multicolinear data) as shown by Cadavez (2009). Hence, the parameters obtained by SUR are characterized by lower standard errors. As the error terms for LMP have the highest correlations with other error terms, the reduction of the standard errors is especially visible in this equation: the SE of the intercept decreases by 44% and the SE of the parameter of HCW decreases by 35%. Thus, modeling the carcass composition of lambs while ignoring the residuals variance-covariance structure results in inefficient estimates (Yahya *et al.* 2008).

Given that the OLS estimates of the unchanged equations are the same in model "E2" and model "E2a", the coefficients of determination of prediction R_p^2 are also the same for these equations. While the removal of the intercept in the equation for SFP increased the R_p^2 by 2.3%, the removal of the intercept and HCW in the equation for KCFP reduced the R_p^2 by 1.8%. When estimating model "E2a" by SUR instead of by OLS, the R_p^2 values for LMP and IFP increase by 1.9% and 1.3%, respectively, while the R_p^2 values for SFP and BP only decrease by 0.8% and 0.3%, respectively. Hence, estimating model "E2a" by SUR improves the average precision of the predictions.

The correlations of the residuals obtained from estimating model "E2a" by OLS and SUR are shown in Table 5.

Table 5: Residuals correlations for model "E2a" estimated by OLS and SUR estimators

	OLS					SUR				
	LMP	SFP	IFP	KCFP	BP	LMP	SFP	IFP	KCFP	BP
LMP	1	-0.708	-0.670	-0.482	-0.189	1	-0.714	-0.673	-0.514	-0.179
SFP		1	0.444	0.538	-0.241		1	0.452	0.574	-0.248
IFP			1	0.323	-0.446			1	0.345	-0.449
KCFP				1	-0.229				1	-0.236
BP					1					1

4. Conclusion

This paper presents a novel approach to simultaneously predict carcass components using the SUR technique and the results are relevant for implementing objective carcass classification systems. Thus, our findings can have a positive effect on the meat industry, since the methodology applied to predict the carcass composition can be integrated into decision support systems in order to use all carcass tissues for the definition of the carcasses' price. The SUR estimator provides the lowest standard errors of the estimated parameters and thus, the highest precision of the estimates. Furthermore, our results revealed that HCW and the E2 measurement are the most relevant predictors of carcass tissues. This study shows that SUR estimates are consistently better than the OLS (equation-by-equation) estimates, since the SUR estimator takes the correlation between the error terms into account . Thus, SUR is a robust methodology for predicting the carcass composition of lambs. In spite of the elegant properties of the SUR estimator, it is an underused multivariate regression technique, especially for predicting carcass composition. Indeed—as far as we know—it has not been used for this purpose before.

References

Cadavez VAP (2009). "Prediction of lean meat proportion of lamb carcasses." *Archiva Zootechnica*, **12**(4), 46–58.

Chatterjee S, Hadi AS, Price B (2000). *Regression analysis by a example*. John Willey & Sons, Inc., New York.

Cortez P, Portelinha M, Cadavez VAP, Rodrigues S, Teixeira A (2006). "Lamb meat quality assessment by support vector machines." *Neural Processing Letters*, **24**, 41–51.

Delfa R, González C, Teixeira A (1996). "Use of cold carcass weight and fat depth measurements to predict carcass composition of rasa Aragonesa lambs." *Small Ruminant Research*, **20**, 267–274.

Ferguson D (2004). "Objective on-line assessment of marbling: a brief review." *Australian Journal of Experimental Agriculture*, **44**(7), 681–686.

Fortin A, Sherestha JNB (1986). "In vivo estimation of carcass meat by ultrasound in ram lambs slaughtered at an live weight of 37 kg." *Animal Production*, **43**, 469–475.

Henningsen A, Hamann JD (2007). "systemfit: A Package for Estimating Systems of Simultaneous Equations in R." *Journal of Statistical Software*, **23**(4), 1–40. URL http://www.jstatsoft.org/v23/i04/.

Hopkins DL (2008). "An industry applicable model for predicting lean meat yield in lamb carcasses." *Australian Journal of Experimental Agriculture*, **48**, 757–761.

Hopkins DL, Ponnampalam EN, Warner RD (2008). "Predicting the composition of lamb carcasses using alternative fat and muscle depth measures." *Meat Science*, **78**, 400–405.

Judge GG, Hill RC, Griffiths WE, Lutkepohl H, Lee TC (1988). *Introduction to the teory and practise of econometrics*. 2 edition. Wiley, New York.

Kongsro J, Roe M, Kvaal K, Aastveit AH, Egelandsdal B (2009). "Prediction of fat, muscle and value in Norwegian lamb carcasses using EUROP classification, carcass shape and length measurements, visible light reflectance and computer tomography (CT)." *Meat Science*, **81**, 102–107.

Lambe NR, Navajas EA, Schofield CP, Fisher AV, Simm G, Roehe R, Bunger L (2008). "The use of various live animal measurements to predict carcass and meat quality in two divergent lamb breeds." *Meat Science*, **80**, 1138–1149.

Montgomery DC (1997). *Design and analysis of experiments*. Fourth edition edition. John Wiley & Sons, New York.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rius-Vilarrasa E, Bünger L, Maltin C, Matthews KR, Roehe R (2009). "Evaluation of video image analysis (via) technology to predict meat yield of sheep carcasses on-line under UK abattoir conditions." *Meat Science*, **82**, 94–100.

Silva SJCR (2001). *Composição das carcaças e dos depósitos internos de gordura de ovelhas de raça Churra da Terra Quente*. Ph.D. thesis, Universidade de Trás-os-Montes e Alto Douro, Vila Real.

Theil H (1971). *Principles of Econometrics*. Wiley, New York.

Wood J (1995). *The influence of carcass composition on meat quality*, chapter Quality and Grading of Carcasses of Meat Animals, pp. 131–155.

Yahya WB, Adebayo SB, Jolayemi ET, Oyejola BA, Sanni OOM (2008). "Effects of non-orthogonality on the efficiency of seemingly unrelated regression (SUR) models." *InterStat Journal*, pp. 1–29. URL http://interstat.statjournals.net/.

Zellner A (1962). "An efficient method of estimating seemingly unrelated regression equations and test for aggregation bias." *Journal of American Statistical Association*, **57**, 348–368.

Affiliation:

Vasco A. P. Cadavez
Mountain Research Center (CIMO)
ESA - Instituto Politécnico de Bragança
Campus de Santa Apolónia, Apartado 1172
5301-855 Bragança, Portugal
E-mail: vcadavez@ipb.pt

Arne Henningsen
Institute of Food and Resource Economics
University of Copenhagen
Rolighedsvej 25
1958 Frederiksberg C, Denmark
E-mail: arne.henningsen@gmail.com
URL: http://www.arne-henningsen.name/