Animal Science 2006, **82**: 919–928
ISSN 1357-7298
DOI: 10.1017/ASC2006104

Prediction of the percentage lean of pig carcasses with a small or a large number of instrumental carcass measurements – an illustration with HGP and Vision

B. Engel[†], E. Lambooij, W. G. Buist, H. Reimert and G. Mateman

Animal Sciences Group, Wageningen UR, PO Box 65, 8200 AB Lelystad, The Netherlands

[†] E-mail: Bas.Engel@wur.nl

Abstract

In this paper we report on the results of a recent dissection experiment in The Netherlands where prediction formulae for the percentage lean meat in pig carcasses with the Hennessy Grading Probe (HGP) and a vision system (from now on referred to as Vision) were determined. Predictions with the HGP were based on one fat and one muscle depth measurement only, while predictions with Vision were based on as many as 115 direct and derived measurements. The data from this dissection experiment were used to illustrate the statistical calculations involved in relation to the number of carcass measurements. Prediction with instruments that gather a large number of measurements per carcass is not covered by the present European Community (EC) regulations. Therefore the calculations were conducted according to new regulations for statistical methodology in pig carcass grading that are expected to be adopted by the EC in the near future. The calculations included consideration of 3 subpopulations (females, entire males and castrated males). The Vision data were also used to show that ordinary regression after selection of a subset of carcass measurements severely under estimates the accuracy of prediction: instruments and associated prediction formulae are seemingly much more accurate than they truly are. When standard regression methods are used for instruments that gather a large number of measurements, there is a considerable risk that measurement instruments will be selected for the wrong reasons. Accuracy of approved instruments may not even comply with the EC-regulations, with poor consequences for harmonization within the EC.

Keywords: *carcass composition, lean, least squares, pigs, prediction, regression analysis.*

Introduction

The Hennessy Grading Probe (HGP; Walstra, 1986) has been used for lean meat prediction of pig carcasses in Dutch slaughterhouses since 1987 and the new prediction formula derived in this paper was an upgrade. The vision system (Vision; VCS 2000; E + V Technology, Oranienburg, Germany) was considered for possible future use in The Netherlands. The new prediction formula for the HGP was derived with ordinary regression analysis, because the number of prediction variables was small: one fat and one muscle depth (measured at the third to fourth from last rib position, 6 cm from the dorsal mid line). For Vision however, the number of prediction variables was large (115 variables, comprising fat and muscle thickness at various positions on the carcass, areas, angles and ratios of measurements). For such a large number of prediction variables, because of high correlations between these variables (multicollinearity), linear regression is no longer a viable option (Montgomery and Peck, 1992; Brown, 1993). Therefore, we used partial least squares (PLS) (Helland, 1988; Martens and Naes, 1989; Brown, 1993). PLS originates from near-infrared spectroscopy, where very large numbers of wavelengths are collected as predictors, for example prediction of protein in 25 samples of flour with some 700 prediction variables (Fearn, 1983). PLS has been applied to regression problems in many different areas of research where the number of prediction variables was (very) large compared with the number of observations. The reasoning behind the PLS method is that underlying the original prediction variables there are a limited number of latent variables that represent the association among the prediction variables and the association between the prediction variables and the response. This idea is reflected in the PLS-algorithm by the calculation of a relatively small set of new variables (often referred to as t-variables) that are linear combinations of the original prediction variables. The number of t-variables used is referred to as the dimension. The final prediction formula is the result of linear regression of the percentage lean meat on the t-variables. PLS is generally found to be quite successful in a variety of applications, although it is mainly motivated by algorithmic considerations without a solid statistical background (therefore considered a soft science application in Stone and Brooks, 1990) and exhibits some undesirable theoretical properties (Butler and

Denham, 2000). As a result of the EUPIGCLASS project (Causeur *et al.*, 2006), PLS was proposed as the new standard for the calculation of prediction formulae for the percentage of lean meat of pig carcasses within the EC.

The aim of this paper is threefold: (1) to present results for prediction with HGP and Vision from the recent Dutch dissection experiment, (2) to illustrate the differences between calculations for measurement instruments that collect a modest number of carcass measurements (HGP) and instruments that collect a large number of carcass measurements for prediction (Vision), and (3) to show that conventional statistical regression methods comprising selection of prediction variables from a large number of carcass measurements can produce seriously misleading results for accuracy of prediction.

Material and methods

Considerations with respect to the sampling procedure

The present European Community regulations (EC; 1985 and 1994) were adopted at a time when a relatively small number of objective measurements were collected per carcass. Standards for accuracy of prediction were established with linear regression or double-regression (a cost saving improvement of linear regression) (Coniffe, 1985; Engel and Walstra, 1991a and b; Causeur and Dhorne, 1998; Causeur, 2005) in mind. Since linear regression performs poorly when a large number of measurements per carcass is used for prediction, Vision could not be evaluated according to the present EC regulations. In the EUPIGCLASS (2000) project the EC regulations were reconsidered in view of the increasing number of measurement instruments that collect large numbers of carcass measurements. PLS was proposed as the new standard method for derivation of a prediction formula for the percentage lean meat (Causeur *et al.*, 2006). The proposed quality criterion for accuracy of prediction was the root mean-square error of prediction (RMSEP). The RMSEP (to be discussed later on) was a natural successor to the root mean square error (RMSE) criterion that features in the present EC-regulations. The RMSEP can be evaluated on the basis of a random sample of carcasses. We decided to adhere to these new rules that are likely to be adopted in the near future. As an obvious reduction of the experimental cost the same dissected carcasses were used for HGP and Vision. Since there was an interest in gender (females, entire males and castrated males), it was decided to collect three separate samples, one for each sex.

For evaluation of accuracy of prediction, random samples of carcasses had to be collected for each sex. However, in pig carcass grading it is extremely difficult, if not impossible, to collect a sample of carcasses that is truly random. This is because practical limitations often do not allow sampling of carcasses from all slaughterhouses in a country or region and throughout the year. Therefore, a proportional sample of 60 carcasses was collected for each sex. A proportional sample mimics a random sample. For each sex, five classes were defined based on the HGP fat depth measurement (at the third to fourth from last rib position, 6 cm from the dorsal mid line) that is presently measured in the slaughter line in Dutch slaughterhouses. Classes were based on HGP fat depth because in The Netherlands this is an important prediction variable that is strongly related to the percentage lean meat. Boundary values for these classes were derived from HGP fat depth values from large samples of carcasses that were collected from five Dutch slaughterhouses prior to the collection of carcasses for dissection. This data about HGP back fat was part of the data that is routinely stored in Dutch slaughterhouses and comprised about 10 000 carcasses of each of the three sexes. The boundary values were chosen as close as possible to the 10, 30, 70 and 90% points for HGP fat depth in the population. Percentile points were estimated for each sex separately. We intended to have six ($=10\%$ of 60) carcasses in the lowest class, 12 ($6+12=18=30\%$ of 60) in the one but lowest class etc. For practical reasons boundary values were chosen in between integer values (in mm). In some cases the intended percentile points had to be slightly shifted and the numbers in the classes were modified accordingly. The boundary values used for boars were 11·5, 13·5, 15·5 and 17·5 (mm), for castrates 13·5, 15·5, 18·5 and 21·5 (mm) and for gilts 11·5, 12·5, 15·5 and 18·5 (mm). The numbers of carcasses for dissection in the HGP fat classes per sex are reproduced in Table 1.

The samples for dissection

Pig carcasses for dissection were selected from four slaughterhouses that were considered representative for the Dutch pig population. Vision was only installed in one of these slaughterhouses; therefore pigs were transported from the other three slaughterhouses to the slaughterhouse where Vision was installed. Approximately 20 pigs were delivered per day and about five were selected on the basis of their HGP fat depth. Carcasses with carcass weight outside the range from 69·7 to 110·9 (kg) (estimated 1 and 99 percentile points respectively) were not included in the sampling procedure to avoid the truly extreme carcasses. The day after slaughter the carcasses were dissected according to the EC-reference method (Walstra and Merkus, 1995). Each carcass was dissected by five butchers.

In the analysis one carcass was considered to be an outlier for Vision on the basis of a very large residual. During the experiment a small number of additional carcasses were selected. Initially the first 180 carcasses from the data file that fitted into the proportional sampling scheme were used in the analyses. Fortunately, we were able to replace the offending carcass by one of the remaining carcasses that fitted into the selection scheme. No other carcasses were replaced or removed. Some summary statistics of the samples are presented in Table 2.

Table 1 *The numbers in the Hennessy Grading Probe (HGP) fat depth classes per sex*

HGP fat class[†]		Boars	Castrates	Gilts
Low HGP fat	1	5	7	10
	2	16	12	8
	3	19	21	23
	4	14	13	14
High HGP fat	5	6	7	5

[†] Boundary values of HGP fat depth classes per sex are presented in the text.

Evaluation of the prediction formulae

The prediction formula for the HGP was derived by ordinary linear regression. For Vision PLS was employed. Both for the HGP and Vision, the accuracy of prediction was expressed in terms of the RMSEP and evaluated by the so-called 'leave-one-out' procedure. RMSEP, leave-one-out and considerations with respect to gender are discussed in the next sections. Two technical remarks before we proceed: (i) throughout this paper PLS was applied to the centred and scaled responses (percentage lean) and prediction variables (Vision measurements) and (ii) variables were centred and scaled afresh each time PLS was applied within the leave-one-out procedure.

Evaluation of accuracy of prediction by RMSEP

The RMSEP is the root of the average squared difference between the actual percentage lean and its prediction. The RMSEP was evaluated by the leave-one-out procedure, in conformance with Causeur *et al*. (2006). Each carcass (say with lean meat percentage LMP) in turn was left out of the sample. A prediction formula was derived from the remaining (n-1) carcasses. With this formula a prediction (say *PLMP*) was derived for the carcass that was omitted. The deletion residual $r = LMP - PLMP$ was calculated. The RMSEP was derived by the following expression, where summation is over the carcasses (n) involved in the calculations:

$$RMSEP = \sqrt{\sum r^2/n}. \quad (1)$$

The sum of squares of the deletion residuals is often referred to as the prediction sum of squares (PRESS) (Montgomery and Peck, 1992). Hence

$$RMSEP = \sqrt{PRESS/n}. \quad (2)$$

Table 2 *Summary statistics for the dissected carcasses: the number of dissected carcasses (n), mean, median, minimum, maximum, standard deviation (mm) and coefficient of variation for the HGP fat and muscle measurements, the lean meat percentage (LMP) and the carcass weight (kg), per sex and overall*

Boars	*n*	mean	median	min.	max.	s.d.	%CV
Fat	60	14·39	14·40	8·40	20·00	2·46	17·1
Muscle	60	50·62	49·80	38·80	68·80	6·22	12·3
LMP	60	56·80	56·81	51·03	63·31	2·60	4·6
Weight	60	81·00	80·45	72·00	94·50	5·43	6·7
Castrates	*n*	mean	median	min.	max.	s.d.	%CV
Fat	60	17·10	17·00	10·40	24·80	3·26	19·0
Muscle	60	57·74	57·40	40·80	77·20	7·98	13·8
LMP	60	55·50	55·64	48·13	66·33	3·49	6·3
Weight	60	89·75	89·45	73·10	110·40	8·26	9·2
Gilts	*n*	mean	median	min.	max.	s.d.	%CV
Fat	60	14·30	14·20	8·80	22·40	2·84	19·9
Muscle	60	56·92	56·85	47·60	79·60	5·78	10·2
LMP	60	58·06	58·08	49·18	67·33	3·43	5·9
Weight	60	88·64	89·10	71·40	107·80	8·40	9·5
Overall	*n*	mean	median	min.	max.	s.d.	%CV
Fat	180	15·27	14·80	8·40	24·80	3·14	20·6
Muscle	180	55·09	54·80	38·80	79·60	7·41	13·5
LMP	180	56·79	56·76	48·13	67·33	3·35	5·9
Weight	180	86·46	86·20	71·40	110·40	8·41	9·7

Obviously, when carcasses are selected, say for extreme values of HGP fat depth, this will affect the RMSEP. Therefore, the RMSEP will be evaluated for a random sample. In that case the RMSEP can be formally regarded as the square root of the expected squared difference between the percentage lean and its prediction for a randomly selected carcass from the pig population.

Why don't we use the predictions and residuals that are derived from the single prediction formula that is based on all n carcasses? The reason is that the latter approach will appear to be too accurate, because the construction and validation of the prediction formula are based on the same data. This makes it (too) easy for the formula to adapt to specific features of the sample that would not be replicated in new samples. Ideally, we would like to have data (usually referred to as the training set) to derive a prediction formula and new data (the validation set) to validate this formula. In the leave-one-out procedure each subset of (n-1) carcasses forms a training set, while the omitted carcass forms a corresponding validation set. When n is not too small, the fact that the RMSEP is derived from prediction formulae based on (n-1) rather than n carcasses can be ignored.

When a formula that was already established has to be validated on a new random sample, the residuals r can be calculated directly from the new data as the differences between the lean meat percentages and their corresponding predictions and the RMSEP again follows from (1).

In past EC regulations the accuracy was expressed in the form of the residual standard deviation (RSD) of linear regression, to be replaced later on by the root mean squared error (RMSE) in order to incorporate prediction error due to possible bias. In linear regression the RMSEP both incorporates the prediction error due to residual variation and the prediction error due to estimation error in the constant and coefficients of the formula. Consequently, the RMSEP will generally be larger than the RSD.

When *leave-one-out* is applied in combination with ordinary linear regression, there is no need to derive the deletion residuals by performing the regression n times. Expressions can be derived (Montgomery and Peck, 1992) that allow the deletion residuals to be obtained from the single regression with all n carcasses. For Vision however, the PLS procedure had to be applied n times. Since PLS involves relatively simple calculations, this offered no problems. We used software from GenStat (2000) and Statistical Analysis Systems Institute (SAS; 1994) (with the same results). In conformance with Causeur *et al*. (2006) but different from GenStat and SAS, in expressions (1) and (2), we used the divisor n, rather than (n-1).

Accuracy of prediction in relation to gender

In order to account for possible effects of gender, the calculations were performed according to the following steps.

1. A prediction formula was derived ignoring the sexes and the RMSEP based on all $n = 180$ carcasses was determined from expression (1). For the HGP, the RSD of the linear regression formula was derived as well.

2. The RMSEP of the overall formula ignoring sexes was also derived for boars, castrates and gilts separately from the deletion residuals for the separate sexes ($n = 60$). For instance, in (1) summation was restricted to the residuals of step 1 of females only. Also, the bias of the overall formula was estimated for each sex from the average of the corresponding deletion residuals for that sex.
3. For each sex a separate prediction formula was derived based on the carcasses for that sex only ($n = 60$) and the RMSEPs were evaluated.
4. For the HGP, main effects (allowing for different constants) and interaction terms (allowing for different coefficients) for gender were added to the regression model and appropriate significance tests (F-tests) for gender effects were performed.
5. For Vision, the prediction formula by PLS in step 1(ignoring gender) is the result of linear regression of the percentage lean on the t-variables. This regression was repeated with extra main effects and interaction terms for gender. F-tests for gender effects were performed.
6. The RMSEP values from steps 2 and 3 were compared to see whether there was any marked gain in the use of separate formulae for the sexes with respect to the accuracy of prediction, as measured by the RMSEP.
7. The bias terms from step 2 were inspected and their impact was considered in relation to the results of the significance tests from steps 4 and 5 and the differences in RMSEP from steps 2 and 3.

Selection of variables

Someone less familiar with PLS might be tempted to improve upon the behaviour of linear regression by selection of a relatively small subset of promising prediction variables. Indeed in our experience this has been the case on a number of occasions in pig classification. Although the idea is quite reasonable, often the statistics employed were not sophisticated enough, with poor consequences for the reported accuracy of prediction. Again, the problem is that variable selection adepts too much to specific features of the data when derivation and validation of the prediction formula are based on the same data. To illustrate this over-optimism, the data set for Vision was randomly (within each sex) split into two parts and one part was used as a training set and the other part as a validation set. The prediction formula was developed with the data from the training set and then applied to the data from the validation set. The training set mimicked the calculations for Vision (but for a smaller set of carcasses). The validation set showed what the predictions would have been for new data. A marked difference in accuracy of the prediction formula between the training and validation set would offer a clear indication of the aforementioned over-optimism. Although variable selection is somewhat alien to the philosophy behind PLS, for completeness sake we attempted to improve upon PLS by variable selection as well, employing an approach similar to Gusnanto *et al.* (2003).

Results

To detect outlying observations we inspected diagnostic plots (reproduced below) of the deletion residuals and the ranges of the prediction variables. No carcasses were removed, except the carcass that was mentioned in section 2 that was replaced prior to the analyses reported here. Although the replaced carcass was clearly an outlier, its replacement hardly affected the results of the analysis.

Effects of gender for HGP

The accuracy of prediction of the HGP, as measured by the RMSEP, for the single overall prediction formula ($n = 180$) ignoring sexes was comparable to the accuracy of prediction of the separate formulae for the sexes ($n = 60$) (Table 3, columns 1 and 3). Therefore, for the HGP, there was no need for the use of separate prediction formulae for the sexes. Any possible bias, as a result of gender effects, was compensated in the RMSEP by the larger sample size of the overall formula.

An estimated relative bias between gilts and castrates of 0·6% (Table 3, column 4, bias $= 0{\cdot}27 + 0{\cdot}32\%$) was found. However, this bias was not significant ($P > 0{\cdot}05$): the F-tests performed gave no indication that gender significantly affected the relationship between percentage lean and HGP fat and muscle depth measurements. Note that gender effects cannot be entirely excluded. However, gender effects were apparently not large enough to be detected by the F-tests performed. It is possible that gender effects would be detected with larger numbers of dissected carcasses. However, the present sample size ($n = 180$) already markedly

Table 3 *Accuracy of lean meat prediction of Hennessy Grading Probe (HGP) (with one fat and one muscle depth measurement, both measured at the third to fourth from last rib position, 6 cm from the dorsal mid line) in relation to gender*

	RMSEP of overall formula (n = 180)†	RSD of overall formula ($n = 180$) and of separate formulae ($n = 60$) for the sexes	RMSEP of separate formulae ($n = 60$) for the sexes‡	Bias of overall formula for the sexes†
Total	2·24	2·21	–	–
Boars	1·94	1·93	1·98	0·05
Castrates	2·37	2·35	2·42	0·27
Gilts	2·37	2·36	2·47	−0·32

† Formula derived from all $n = 180$ carcasses, but RMSEP and bias for separate sexes derived from the 60 deletion residuals for each sex only.
‡ Formulae per sex derived from the $n = 60$ dissected carcasses for that sex only.

exceeded the minimal sample size of $n = 120$ as required for a dissection experiment by the EC regulations. No power calculations were performed, but there is little doubt that substantially larger sample sizes would be needed to detect any effects of gender and still larger sample sizes to improve upon prediction by inclusion of gender effects. With the present sample size, the standard errors on the estimated gender effects were too large for gender to improve upon the accuracy of prediction of the HGP.

In the separate regressions per sex the *RSD* for boars (RSD = 1·93) was somewhat lower than for castrates and gilts (RSD = 2·35 and 2·36 respectively) (column 2 of Table 3). However, when compared pairwise by F-tests, the RSDs did not significantly ($P > 0{\cdot}05$) differ.

The prediction formula for the HGP in The Netherlands

The prediction formula for the HGP, as obtained by linear regression, ignoring the sexes, reads as follows:

$$PLMP = 60{\cdot}85 - 0{\cdot}745^{*} \times Fat + 0{\cdot}133 \times Muscle, \quad n = 180, \quad RMSEP = 2{\cdot}24. \tag{3}$$

Here, *PLMP* = the predicted lean meat percentage and *Fat* and *Muscle* are the fat and muscle depths as measured by the HGP at the third to fourth from last rib position, 6 cm from the dorsal mid line. The *RMSEP* was below the upper bound of 2·5% that is specified in the EC regulations.

The prediction formula for the HGP that is presently used in Dutch slaughterhouses was derived from a dissection experiment conducted in 1990 (Engel and Walstra, 1993). Predictions with this formula are derived with the same fat and muscle depth measurements as in expression (3), but are weighted means of separate predictions for gilts and castrates (Engel and Walstra, 1993). The weights, that are basically probabilities that the carcass is either from a gilt or a castrate, depend on the HGP fat and muscle depth measurements. A linear approximation to this non-linear formula (Engel, 1991) closely resembles the new HGP formula from expression (3). Additional calculations (not shown), employing methodology similar to Font i Furnols *et al*. (2004), confirmed that the currently used formula and the new HGP formula from (3) were surprisingly similar and that the accuracy of the presently used prediction formula was still in compliance with the EC-rules. We concluded that any changes that may have occurred in the Dutch pig population after 1990 apparently did not affect the relationship between the percentage lean and the HGP measurements to any great extent. Figure 1 shows a plot of the deletion residuals against the predicted values for the HGP.

Effects of gender for Vision

An overall prediction formula (ignoring gender) was obtained by PLS. This prediction formula, involving all 115 prediction variables, is presented in Appendix A. The results for PLS (Table 4) were based on the use of 4 t-variables (dimension = 4). This choice was based on the profile of the RMSEP for increasing dimension and was made automatically in SAS and in GenStat (Osten, 1988). On the

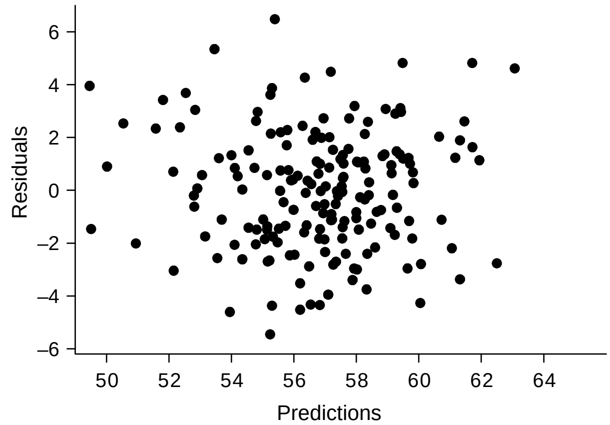

Figure 1 Deletion residuals against predictions for Hennessy Grading Probe (HGP).

basis of the RMSEP (Table 4, columns 1 and 2) there was no need for separate formulae for the sexes.

A sizeable estimated relative bias between castrates and gilts of 1% (Table 4, column 3; bias = 0·54 + 0·43%) was found. When main effects and interaction terms for gender were added to the regression of percentage lean on the t-variables, the interaction terms were not significant ($P = 0{\cdot}31$). However, main effects for gender were significant ($P = 0{\cdot}05$). Pairwise comparisons indicated that castrates and gilts differed significantly ($P = 0{\cdot}01$), but castrates and boars and gilts and boars did not ($P = 0{\cdot}36$ and $P = 0{\cdot}15$ respectively). The estimated difference between castrates and gilts, from the regression model with main effects for sexes, was also equal to1 (se = 0·4) %.

Although effects of gender were found to be statistically significant, the RSD of the regression of percentage lean on the t-variables was only marginally decreased from 1·98 to 1·95% when gender effects were included. This confirmed that gender effects did not add to the quality of prediction to any great extent: the RMSEP values in Table 4 showed that predictions from separate formulae for the sexes were less accurate than predictions from a single overall prediction formula. As a final check we added two dummy variables for the sexes to the 115 predictors from Vision in the PLS analysis of lean meat percentage, thus effectively allowing for different constants for the sexes, but equal coefficients for the 115 prediction variables. The RMSEP was 2·15 and only marginally lower than the value 2·19 (Table 4) that

Table 4 *Accuracy of lean meat prediction of Vision (with 115 carcass measurements, see Appendix A) in relation to gender*

	RMSEP of overall formula ($n = 180$)†	RMSEP of separate formulae ($n = 60$) for the sexes‡	Bias of overall formula for the sexes†
Total	2·19	–	–
Boars	2·30	2·40	−0·05
Castrates	2·11	2·47	−0·43
Gilts	2·16	2·58	0·54

† Formula derived from all $n = 180$ carcasses, but RMSEP and bias for separate sexes derived from the 60 deletion residuals for each sex only.
‡ Formulae per sex derived from the $n = 60$ carcasses for that sex only.

was found when the sexes were ignored. Hence, we concluded (again) that there is no need in The Netherlands to allow for gender effects in the prediction of the lean meat percentage by Vision, neither by the use of completely separate formulae, nor by the use of separate constants (and common coefficients) for gender.

Validation of RMSEP and leave one out for Vision

In this section our aim was not to re-estimate the *RMSEP*, but to add confidence to the reliability of the PLS method in combination with the *leave-one-out* method. The Vision data were randomly divided into two parts (randomly within each sex). One part (the training set, $n = 90$) was used to develop a prediction formula by PLS (ignoring sexes) employing all 115 Vision variables. The other part (the validation set, $n = 90$) was used to see how this formula performed on new data. Note that the RMSEPs will be larger than the RMSEPs reported for the complete data set in Table 4, because the training and validation sets were only half as large ($n = 90$) as the total data set ($n = 180$). The RMSEP of the prediction formula obtained by PLS was derived from the training set by leave-one-out, in the same way as described before, employing (1) with $n = 90$ deletion residuals. The same formula was applied to the validation set and the RMSEP was obtained from the residuals that were calculated as the differences between the lean meat percentages and their predictions. The random division into two sets was performed 50 times. We found an average RMSEP of 2·45 (RMSEP varied from 2·1 to 2·9) for the training set and an average RMSEP of 2·35 (RMSEP varied from 2·0 to 2·8) for the validation set. So, the RMSEP as calculated by leave-one-out and PLS with all prediction variables included offers a reliable impression of accuracy, since the accuracy for the training set applies to new data (represented by the validation set) as well.

Variable selection for PLS on the Vision data

We proceeded similar to a method recently suggested by Gusnanto *et al*. (2003). In this *forward* procedure, at a certain stage we had a subset of prediction variables that were already chosen. Each of the remaining variables in turn was temporarily added to the current subset and the RMSEP of PLS was determined with leave-one-out. The variable that corresponded to the lowest RMSEP was added to the subset. This procedure was repeated with the new and larger subset and stopped when a pre-chosen maximum number of prediction variables (5, 10 or 20) was reached. Each time PLS was used, the dimension was chosen afresh. The procedure was applied 40 times on a training set of 90 carcasses and a validation set of 90 carcasses. The division into the two sets was performed randomly within the sexes. The results for the RMSEP are shown in Table 5. The RMSEPs with PLS with all Vision measurements included for the training and validation sets were similar to those reported before (on average around 2·4). We concluded that no improvement was obtained with variable selection. For the validation set the RMSEP was always markedly higher than for the training set. RMSEP values that are too optimistic can be avoided by including the selection process within a leave-one-out procedure. This will be rather computer intensive since the whole selection approach, that already involves leave-one-out in its different selection steps, will have to be performed 180 times (leaving out each observation in turn).

Table 5 *Results for root mean-square error of prediction (RMSEP) with partial least squares (PLS) and variable selection (cross validation with 40 random splits) for the Vision data*

No. of prediction variables	Average RMSEP training set (minimum; maximum)	Average RMSEP validation set (minimum; maximum)
5	2·18 (1·92; 2·44)	2·71 (2·41; 3·16)
10	1·91 (1·62; 2·10)	2·59 (2·30; 2·98)
20	1·75 (1·42; 2·04)	2·53 (2·13; 3·03)

Although attempts have been made in the literature to combine PLS with variable selection, this seems to be rather counter-intuitive, since PLS automatically down-weighs the less important prediction variables. We conjecture that in lean meat prediction, where the relationship between the carcass measurements (including derived variables) is less obvious than between neighbouring wavelengths in near-infra red measurements (the problem that originated PLS), there is no appreciable gain in variable selection. For that reason we recommend the use of the prediction formula with all 115 Vision measurements. This prediction formula is reproduced in appendix A, together with a list of the variables involved. A plot of the deletion residuals against the predictions for Vision is shown in Fig. 2.

Stepwise linear regression for the Vision data

Accuracy of prediction under variable selection in regression can be severely under-estimated. To illustrate this phenomenon, again we randomly divided the Vision data into two parts (randomly within each sex). One part (the training set, $n = 90$) was used to develop a prediction formula by stepwise regression. In stepwise regression prediction variables are included or left out of the regression according to the F-ratio that compares the model with and without the prediction variable. The F-ratio's to include or leave out were both set equal to 1. For details see GenStat (2000), Section 3·2.5. Variable selection was performed first and the selected variables were subsequently used in the calculation of the

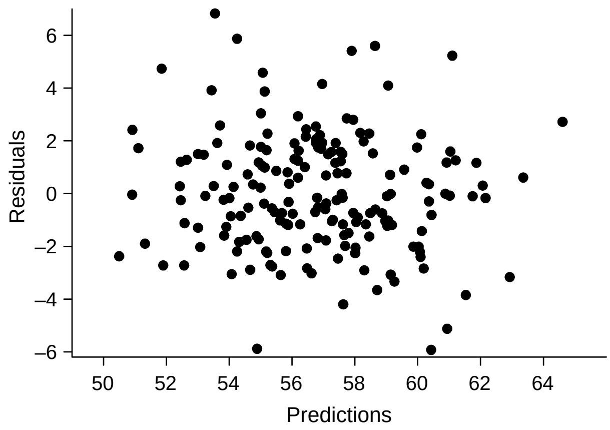

Figure 2 Deletion residuals against predictions for Vision.

RMSEP by the leave-one-out procedure. The same prediction formula was applied to the validation set and the RMSEP was evaluated again from the residuals calculated as the differences between the lean meat percentages and their predictions. This process was repeated 50 times for a maximum of 5, 10 or 20 Vision variables in the regression. The results are summarized in Table 6. Clearly, the reduction in RMSEP achieved by stepwise regression for the training set was lost when the same formula was applied to the validation set. In other words, the accuracy of the regression formula after variable selection did not extend to new data and therefore did not offer a reliable impression of the performance of the prediction formula in practise.

We can obtain a better impression of the accuracy of linear regression after variable selection by including the selection procedure in the leave-one-out method. Each observation in turn was left out and stepwise regression was applied to the remaining 179 observations. A prediction was made for the omitted observation and the deletion residual was calculated as the difference between the observation and its prediction. From the 180 deletion residuals the RMSEP was calculated according to expression (1) ($n = 180$). For a maximum of 5, 10 and 20 selected Vision variables, the RMSEPs were 2·56, 2·37 and 2·32% respectively. These RMSEP values may be compared with the RMSEP of 2·19 (Table 4) obtained with PLS. We concluded that stepwise regression did not improve upon PLS.

Discussion

The accuracy of prediction was evaluated by means of the RMSEP and calculated by the leave-one-out procedure. Calculation of the RMSEP from a random sample is straightforward. Random samples, however, are often hard to obtain, for instance because sampling over all slaughterhouses in a region and throughout the year for all relevant seasons is too time consuming and costly. Therefore a proportional sample was taken. A proportional sample mimics a random sample. For the proportional sample we have chosen a classification on the basis of an important prediction variable that was measured in the slaughterline. Here, the classification was based on a fat depth measured by the HGP. From large data sets that are stored by Dutch slaughterhouses, five classes and proportions of animals in these classes were established and by selection on the HGP fat measurement the same proportions were imposed on the sample. Obviously, the HGP fat depth measurement will not be used as a predictor in conjunction with Vision measurements. In Engel *et al.* (2003) it is shown that selection on the basis of a variable that will not be included as a prediction variable in the final prediction formula may result in poor predictions. However, this result does not apply here, because samples are proportional and for all practical purposes can be regarded as random samples.

Table 6 *Results for root mean-square error of prediction (RMSEP) with stepwise regression (cross validation with 50 random splits) for the Vision data*

No. of prediction variables	Average *RMSEP* training set (minimum; maximum)	Average *RMSEP* validation set (minimum; maximum)
5	2·31 (1·89; 2·87)	2·67 (2·20; 3·17)
10	2·12 (1·66; 2·79)	2·54 (2·21; 2·99)
20	1·88 (1·46; 2·31)	2·60 (2·30; 3·03)

In principle it should be possible to combine the PLS method with the cost saving double regression method (Engel and Walstra, 1991a and b; Causeur and Dhorne, 1998; Causeur, 2005). However, the statistical evaluation of such a combination was beyond the scope of the present experiment and we preferred the transparency of three separate samples of carcasses for gilts, for boars and for castrates that were all dissected according to the EC-reference method.

Selection on extreme values of the prediction variables generally improves upon the accuracy of estimation of constants and coefficients in regression. However, in carcass grading the gain is likely to be small (Engel *et al.*, 2003). Since evaluation of the RMSEP of PLS for a non-random sample is not straightforward, we preferred to take random samples (or rather proportional samples).

It was shown how the accuracy of prediction of one single overall formula could be compared with the accuracy of separate prediction formulae for the 3 sexes. For HGP we found no significant effects of gender. For Vision we found that, although effects of gender were significant, accuracy of prediction was not improved by use of separate formulae for the sexes.

We conjectured that there is no appreciable gain in accuracy of prediction, as expressed by RMSEP, by variable selection in combination with PLS. Of course in a different situation when some measurements are relatively expensive (or time consuming) to collect, a compromise may have to be struck between cost and accuracy. In that case *a priori* a limited number of sets of prediction variables of increasing cost could be distinguished and PLS could be applied for each set.

We showed that accuracy of prediction of ordinary linear regression in combination with variable selection can be far too optimistic when the selection of the prediction variables is not properly accounted for. Prior to approval in Brussels, such misleading calculations may adversely affect comparisons between different (competing) instruments. This may lead to the wrong choice of instrument and acceptance of an instrument and prediction formula that in reality may not be in compliance with EC regulations for prediction accuracy at all. Proper RMSEP values can be obtained by including the selection process in a leave-one-out procedure. However, on the basis of our results for the Vision data, we prefer the use of PLS with the full set of carcass measurements (including reasonably motivated derived variables).

Acknowledgements

The authors would like to thank the director and staff of HMG, Stork RMS and E + V for managing the butchers and facilities at the slaughterhouse and C. Richter and S. Lescouhier (University of Ghent) for technical support. This study was financed by the Dutch Product Board for Livestock, Meat and Eggs (PVE).

References

Brown, P. J. 1993. *Measurement, regression and calibration*. Clarendon Press, Oxford.

Butler, N. A. and Denham, M. C. 2000. The peculiar shrinkage properties of partial least squares regression. *Journal of the Royal Statistical Society B* **62:** 585-593.

Causeur, D. J. 2005. Optimal sampling from concomitant variables for regression problems. *Journal of Statistical Planning and Inference* **128:** 289-301.

Causeur, D. J. and Dhorne, T. J. 1998. Finite sample properties of a multivariate extension of double regression. *Biometrics* **54:** 1591-1601.

Causeur, D., Daumas, G., Dhorne, T., Engel, B., Font i Furnols, M. and Højsgaard, S. 2006. *Statistical handbook for assessing pig classification methods* (In press).

Coniffe, D. 1985. Estimating regression equations with common explanatory variables but unequal numbers of observations. *Journal of Econometrics* **27:** 179-196.

Engel, B. 1991. *Statistical analysis of the Dutch dissection data concerning harmonisation of methods for grading pig carcasses in the community*. Animal Sciences Group, Wageningen UR, Lelystad, The Netherlands.

Engel, B. and Walstra, P. 1991a. Increasing precision or reducing expense in regression experiments by using information from a concomitant variable. *Biometrics* **47:** 13-20.

Engel, B. and Walstra, P. 1991b. A simple method to increase precision or reduce expense in regression experiments to predict the proportion of lean meat of carcasses. *Animal Production* **53:** 353-359.

Engel, B. and Walstra, P. 1993b. Accounting for subpopulations in prediction of the proportion of lean meat of pig carcasses. *Animal Production* **57:** 147-152.

Engel, B., Buist, W. G., Walstra, P., Olsen, E. and Daumas, G. 2003. Accuracy of prediction of percentage lean meat and authorization of carcass measurement instruments: adverse effects of incorrect sampling of carcasses in pig classification. *Animal Science* **76:** 199-209.

EUPIGCLASS. 2000. *Standardisation of pig carcasses classification in the EU through improved statistical procedures and new technological developments*. Program funded under the Growth Programme. www.cordis.lu/en/home.html

European Community. 1985. *EC regulation no. 2967/85, laying down detailed rules for the application of the community scale for grading pig carcasses*.

European Community. 1994. *EC regulation no. 3127/94, amending regulation (EC) no. 2967/85, laying down detailed rules for the application of the community scale for grading pig carcasses*.

Fearn, T. 1983. A misuse of ridge regression in the calibration of a near infrared reflectance instrument. *Applied Statistics* **32:** 73-79.

Font i Furnols, M., Engel, B. and Gispert, M. 2004. Validation of the Spanish equation to predict the lean meat percentage of pig carcasses with the Fat-O-Meater. *Spanish Journal of Agricultural Research* **2:** 545-549.

GenStat Committee. 2000. *The guide to GenStat.* (ed. R. Payne), VSN, International Ltd, Oxford.

Gusnanto, A., Pawitan, Y., Huang, J. and Lane, B. 2003. Variable selection in random calibration of near-infrared instruments: ridge regression and partial least squares regression settings. *Journal of Chemometrics* **17:** 174-185.

Helland, I. S. 1988. On the structure of partial least squares regression. *Communications in Statistics and Simulation* **17:** 581-607.

Martens, H. and Naes, T. 1989. *Multivariate calibration*. John Wiley, Chichester.

Montgomery, D. G. and Peck, E. 1992. *Introduction to linear regression analysis*, *Second Edition*. John Wiley, New York.

Osten, D. W. 1988. Selection of optimal regression models via cross-valiadation. *Journal of Chemometrics* **2:** 79-92.

Statistical Analysis Systems Institute. 1994. *User's guide version 6*, *Fourth Edition*. SAS Institute Inc., Cary, NC.

Stone, M. and Brooks, R. J. 1990. Continuum regression: cross validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (with discussion). *Journal of the Royal Statistical Society B* **52:** 237-269.

Walstra, P. 1986 *Assessment of the regression formula for estimation of the lean meat percentage by HGP-measurements in the Netherlands*. EC working paper, Brussels, VI/4849/86

Walstra, P. and Merkus, G. S. *Procedure for assessment of the lean meat percentage as a consequence of the new EC reference dissection method in pig carcass classification*. Animal Sciences Group, Wageningen UR, Lelystad, The Netherlands.

(*Received* 19 *December* 2005—*Accepted* 10 *August* 2006)

Appendix A *The prediction formula for Vision*

Constant (intercept) = 40·34501933.

Coefficients (slopes) of Vision measurements(x-variables):

x1	−0·02578308	x2	0·01819554	x3	−0·01880347
x4	0·01336851	x5	0·03402026	x6	−0·00431013
x7	0·01399586	x8	−0·01685383	x9	0·03029107
x10	0·00001618	x11	0·00003510	x12	0·00017285
x13	−0·00006323	x14	−0·00007814	x15	0·03852565
x16	0·03703442	x17	0·04915633	x18	0·16952262
x19	−0·07474948	x20	−0·00000064	x21	−0·00000045
x22	−0·00000217	x23	0·00007137	x24	0·00001311
x25	0·00005955	x26	0·00021689	x27	0·00003631
x28	0·00003996	x29	0·00010888	x30	0·00000010
x31	−0·00002481	x32	−0·00003633	x33	−0·00000496
x34	−0·00000991	x35	−0·00000980	x36	−0·00001532
x37	−0·00000254	x38	−0·00000886	x39	0·00064604
x40	−0·00057968	x41	0·00012513	x42	−0·00002055
x43	−0·00000035	x44	−0·00001008	x45	−0·00017158
x46	0·00019477	x47	−0·00329811	x48	0·00018512
x49	0·02520693	x50	−0·02466706	x51	−0·00286098
x52	−0·01983356	x53	−0·00012369	x54	−0·04044495
x55	0·00069282	x56	0·00025689	x57	0·00403378
x58	0·00028090	x59	10·79585085	x60	−0·65106186
x61	−0·29583150	x62	0·16410084	x63	0·06306376
x64	−2·22708974	x65	0·09886817	x66	0·55302524
x67	−1·38136772	x68	−12·51618138	x69	−2·65248635
x70	9·08361318	x71	−0·00003025	x72	−0·00043013
x73	−0·00083619	x74	−0·00013210	x75	0·00002221
x76	−0·00007579	x77	−0·00031534	x78	−0·00001867
x79	−0·00003067	x80	0·00002421	x81	−0·00026088
x82	0·00067359	x83	0·00024535	x84	−0·00000847
x85	0·00003206	x86	0·00007141	x87	0·00006951
x88	−0·04465410	x89	0·01774709	x90	−0·03618843
x91	−0·02700865	x92	−0·02221780	x93	−0·03644020
x94	−0·01766434	x95	0·00516267	x96	−0·02958650
x97	−0·01189834	x98	0·00184860	x99	0·00894691
x100	0·01586154	x101	−0·01265988	x102	−0·00912249
x103	0·02734395	x104	0·13333009	x105	0·18105960
x106	−0·06387375	x107	0·03471900	x108	0·21910356
x109	−0·12423421	x110	0·09408101	x111	0·01363443
x112	0·06193271	x113	−0·02802852	x114	0·02757767
x115	0·00035641				

The original (German) names of the *x*-variables (details available from the second author) are:

1 Schinkenwinkel
2 Schinkenbreite
3 Taillenbreite
4 Schinkenbreite_Innen
5 Schinkenbreite_Aussen
6 Schinkenlenge
7 Schinkenbreite_Mittel
8 Rueckenbreite
9 Mittlschinkenbreite1
10 Schinkenflaeche_Ges
11 Schinkenflaeche_Aussen
12 Schinkenflaeche_Innen_1
13 Schinkenflaeche_Innen_2
14 Hesseflaeche
15 Wi_SPG
16 Wi_SPG_SCHG
17 DWi_WS_bA_KB
18 DWi_WS_bA_SPG
19 DWi_WS_vB_SPG
20 FL_ges
21 FL_ges_O
22 FL_ges_U
23 FL_b_19
24 FL_b_19_U
25 FL_v_19_b_14
26 FL_v_19_b_14_U
27 FL_v_14_b_13
28 FL_v_14_b_13_O
29 FL_v_14_b_13_U
30 FL_v_13_b_23
31 FL_v_23_b_56
32 FL_v_56
33 FL_v_13_b_PS2_ges
34 FL_v_13_b_PS2_O
35 FL_v_13_b_PS2_U
36 FL_SPG_WS
37 FL_WS_OK
38 FL_WS_UK
39 FL_SPG_WS_273
40 FL_WS_OK_273
41 FL_WS_UK_273
42 FL_SPG_WS_3end
43 FL_WS_OK_3end
44 FL_SPG_UK_3end
45 EF_SP_vP
46 DifX_SP_vP
47 DifY_SP_vP
48 EF_PS1_PS2
49 EF_SP_PS1
50 EF_vP_PS2
51 BREITE
52 BREITE_13_PS2
53 LEN_WS
54 LEN_WS_273
55 LEN_WS_3end
56 LEN_WS_SPG
57 LEN_WS_SPG_273
58 LEN_WS_SPG_3end
59 V_fl_ges_O_U
60 V_fl_vbD_OzU
61 V_fl_DbPS2_OzU
62 V_LAE_bR
63 V_fl_WS_OKzSPG
64 V_fl_WS_OKzSPG_273
65 V_fl_WS_OKzSPG_3end
66 V_fl_WS_SPG_OzU
67 V_fl_WS_SPG_OzU_273
68 V_len_WS_SPG
69 V_len_WS_SPG_273
70 V_len_WS_SPG_3end
71 Spfl_ges
72 Spfl_1
73 Spfl_2
74 Spfl_3
75 Spfl_4
76 Spfl_5
77 Spfl_1_2
78 Spfl_3_4
79 Spfl_1_4
80 FL_WS_SP_ges
81 FL_WS_SP_1
82 FL_WS_SP_2
83 FL_WS_SP_3
84 FL_WS_SP_4
85 FL_WS_SP_5
86 FL_WS_SP_1_2
87 FL_WS_SP_3_4
88 MINFETT
89 MAXFETT
90 MAXFETT_5
91 SPM_ges
92 SPM_1
93 SPM_2
94 SPM_3
95 SPM_4
96 SPM_5
97 SPM_1_4
98 HAM_br_13
99 HAM_br_38
100 KOT_br_38
101 KOT_br_13
102 MTL_KOT_br_fl_5
103 MTL_fl_br_fl_5
104 V_Spfl_34z12
105 V_fl_WS_SP_34z12
106 V_Wsfl_Spfl_1
107 V_Wsfl_Spfl_2
108 V_Wsfl_Spfl_3
109 V_Wsfl_Spfl_4
110 V_Wsfl_Spfl_5
111 V_Wsfl_Spfl_1_2
112 V_Wsfl_Spfl_3_4
113 SPECK
114 KOT
115 LAENGE