Running headline: Prediction precision - lean meat percentage in pork

Title: Prediction precision for lean meat percentage in Norwegian pig carcasses using 'Hennessy grading probe 7'. Evaluation of methods emphasized at exploiting additional information from Computed Tomography.

Lars Erik Gangsei
a: Animalia - Norwegian Meat and Poultry Research Centre
P.O. 396 - Økern
N-0513 OSLO, Norway
lars.erik.gangsei@animalia.no
b: Norwegian University of Life Sciences
Dep. of Chemistry, Biotechnology and Food Science
P.O. Box 5003
N-1432 Ås, Norway
+47 950 61 231

Jørgen Kongsro
Norsvin SA
P.O. Box 504
N-2304 Hamar, Norway
jorgen.kongsro@norsvin.no
+47 915 62 852

Eli Vibeke Olsen
Danish Meat Research Institute (DMRI)
Gregersensvej 9
DK-2630 Taastrup, Denmark
evo@teknologisk.dk
+45 72 20 25 98

Morten Røe
Animalia - Norwegian Meat and Poultry Research Centre
P.O. 396 - Økern
N-0513 OSLO, Norway
morten.roe@animalia.no
+47 915 47 910

Ole Alvseike
Animalia - Norwegian Meat and Poultry Research Centre
P.O. 396 - Økern
N-0513 OSLO, Norway
ole.alvseike@animalia.no
+47 23 05 97 98

Solve Sæbø
Norwegian University of Life Sciences
Dep. of Chemistry, Biotechnology and Food Science
P.O. Box 5003
N-1432 Ås, Norway
solve.sabo@nmbu.no
+47 67 23 25 61

Corresponding author:
Lars Erik Gangsei
a: Animalia - Norwegian Meat and Poultry Research Centre
P.O. 396 - Økern
N-0513 OSLO, Norway
lars.erik.gangsei@animalia.no
b: Norwegian University of Life Sciences
Dep. of Chemistry, Biotechnology and Food Science
P.O. Box 5003
N-1432 Ås, Norway
+47 950 61 231

This work was supported by the Research Council of Norway under Grant ['PigComp' (project number 225294)].

All authors declare no conflict of interest.

Key Words: empirical Bayes; sufficient sample size; surrogate variable;

ABSTRACT

The present study aims at improving the prediction of lean meat percentage (LMP) for pig carcasses based on on-line measurements from the slaughterhouses using the 'Hennessy Grading Probe 7' (HGP7) and auxiliary information such as gender and breed. The prediction performance is evaluated using an empirical Bayes method capable of utilizing information from a surrogate variable, i.e. LMP from computed tomography (CT).

HGP7 measures thicknesses of fat and meat layers. The HGP7 measurements of subcutaneous fat, sirloin height and interior fat layer should be included as predictor variables together with gender. For efficiency at the slaughter–line gender might be omitted.

The empirical Bayes method improved prediction precision only marginally compared with the standard ordinary least squares method when applied to the full set of data. However, simulations show that the empirical Bayes method enables a considerable reduction of the data sample size without appreciable loss of prediction precision.

1 Introduction

The lean meat percentage (LMP) is used as the primary classification variable for the quality of pig carcasses in the European Union (EU). LMP is defined as the proportion of weight of lean meat to the total weight of the carcass (Commission of the European Communities, 2008). Measuring the LMP is typically done using ultrasound or optical probes. Norwegian slaughterhouses use the optical probe 'Hennessy Grading Probe 7' (HGP7).

The data obtained from optical probes or ultrasound do not provide a direct measurement of LMP, but are used as predictor variables in a regression equation of which the sole purpose is LMP prediction. Due to rapid evolution in the pig population, and to maintain high public confidence in on–line predictions of LMP, the prediction equation for LMP is updated on a regular basis. In Norway this is done approximately every fifth year, including the years 2008 and 2013.

The parameters in the regression equation are estimated using a training sample of pig carcasses where LMP is measured by at least one independent reference method. Until 1 January 2009, manual dissection (Walstra and Merkus, 1995) was the only method approved as an official reference method by the EU–authorities. From this date, LMP estimates from computerized tomography (CT) was approved as an offcial reference method for LMP (Commission of the European Communities, 2008). Throughout this text, LMP–MD and LMP–CT will denote LMP obtained by manual dissection and CT respectively.

EU regulations imposes requirements to prediction precision for LMP stating that: '*Grading methods shall be authorised only if the root mean squared error of prediction (RMSEP), computed by a full cross-validation technique or by a test set validation on a representative sample of at least 60 carcases, is less than 2,5. In addition, any outliers shall be included in the calculation of RMSEP*' (Commission of the European Communities, 2008).

Most countries use an equation which predicts LMP on a manual dissection scale. However, CT scanning is cheaper than manual dissection. Furthermore, the LMP–CT estimates are believed to be more consistent, they are possible to replicate and they have a very high correlation with LMP–MD (Vester-Christensen et al., 2009; Judas et al., 2007). Therefore many countries use CT–dissection as their main reference method (Daumas and Monziols, 2011), usually supported by manual dissection for a subset of the carcasses.

The choice of method for estimating regression parameters is a trade–off between com-

plexity and accuracy. We want a method obtaining regression parameters providing small prediction errors for the whole population of pig–carcasses. On the other hand, we want the method to be simple, and preferably to yield easily interpretable results. In this study, we examine a totally data–driven empirical Bayes method (Gangsei et al., 2016) capable of utilizing information from the additional observations of LMP–CT to improve regression estimates for LMP–MD.

In addition, to find a suitable regression equation for LMP–MD based on the predictor variables in question, this study aims at examining the relationship between sample sizes of manual dissected and CT scanned carcasses and prediction precision for LMP–MD.

The program 'R' (R Core Team, 2014), including the package 'bestglm' (McLeod and Xu, 2011), is used for statistical computing. The R code used in this study can be provided upon request.

2 Material and Methods

2.1 Data

The data consist of CT scans of 465 Norwegian half pig carcasses of which 229 were scanned in 2008 and 236 in 2013. The pigs were slaughtered at two different commercial Norwegian abattoirs. The CT–scanning was performed by Danish Meat Research Institute (DMRI). The carcasses were prepared, scanned and weighed in line with the description in Vester-Christensen et al. (2009). Based on these CT–data DMRI predicted LMP–CT for all carcasses using the method described in (Vester-Christensen et al., 2009).

The carcasses were stratified to 4×4 different classes based on weight and on–line slaughterhouse HGP7 measurements to ensure data sampling across the entire range of carcasses. The HGP7-probe penetrates the rind, and measures the thickness of subcutaneous fat (denoted 'Fat'), sirloin ('Meat') and the interior fat layer ('Totif') under the peritoneum. The measurements are done at two specified locations, one behind the last rib and 8 cm from the spine midline, and the other 12 cm further forward on the carcass, 6 cm from the spine midline. For all carcasses on–line data from the HGP7, weight, gender (castrates or females) and breed were registered.

A sample of the CT scanned carcasses were transferred to Animalia's pilot plant in Oslo

where manual dissection was carried out as described by Walstra and Merkus (1995), and the LMP–MD was calculated using the formula according to Commission of the European Communities (2006). A total of 86 carcasses were manually dissected, of which 66 were dissected in 2008 and 20 in 2013.

The maximum number of primary predictor variables was seven, of which four were continuous variables, Fat, Meat, Totif (from HGP7-measurement) and 'Weight'. Note that Fat was the mean of the two fat–measurements, since they were highly correlated ($\rho = .824$), whereas the measurements for Meat and Totif were single measurements from the foremost measurement point. The correlations between the four continuous predictor variables were low to moderate, i.e. in the range -0.49 until 0.48. In accordance with Gangsei et al. (2016) continuous predictor variables, and the response variables, were centred prior to the analysis.

There were three factorial variables. 'Year' had two levels, 2008 (n=229) and 2013 (n=236). The variable 'Gender' had two levels, females (n=228) and castrates (n=237). 'Breed' had three levels, Hampshire (LYHH) (n=69), Duroc (LYLD) (n=243) and Norhybrid (LYLY) (n=153), representing three different hybrids used in Norway. The maternal line for all three hybrids were crossings between Norwegian Landrace and Yorkshire (denoted LY). The paternal lines were pure Hampshire (HH), crossing between Norwegian Landrace and Duroc (LD) and crossing between Norwegian Landrace and Yorkshire (LY) respectively. In the appendix (A.1) summary statistics for LMP–MD, LMP–CT and the continuous predictor variables crossed over the factor variables Year and Breed are presented.

2.2 Model

In this paper we use notations in line with Gangsei et al. (2016). LMP–MD is a vector of length $n_2 = 86$ denoted $\boldsymbol{y}_{MD}$ (corresponding to $\boldsymbol{y}_2$ in Gangsei et al. (2016)), and LMP–CT is a vector of length $n_1 = 465$, denoted $\boldsymbol{y}_{CT}$ (corresponding to $\boldsymbol{y}_1$). Then the $n_1 \times 2$ matrix $\boldsymbol{Y}_1$ contains a substantial number of unobserved LMP–MD's.

The matrix $\boldsymbol{X}_1$ is an $n_1 \times p$ matrix of predictor variables where the first column is the unit vector. The next $p-1$ columns are the predictor variables. In some parts of the analysis second–order interaction terms are included as extra predictor variables in $\boldsymbol{X}_1$.

The formal model is given in (1), where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ denotes the ith row of $\boldsymbol{X}_1$ and $\boldsymbol{Y}_1$ respectively. Note the matrix-form of the regression–parameter ($\boldsymbol{\beta} = [\boldsymbol{\beta}_{CT} \ \boldsymbol{\beta}_{MD}]$) and that

the error–terms for each observation might be correlated with covariance matrix $\boldsymbol{\Sigma}$.

$$\boldsymbol{y}_i^T \overset{i.i.d.}{\sim} N_2\left(\boldsymbol{\beta}^T \boldsymbol{x}_i^T,\ \boldsymbol{\Sigma}\right),\ \ i = 1, \ldots, n_1 \tag{1}$$

Gangsei et al. (2016) deals with this model in detail under an empirical Bayes inference. Empirical Bayes methods uses the data to fit the prior distribution, including the hyperparametes, i.e. the parameters used in the prior distribution. In a strict Bayesian sense this is 'cheating'. However, in a number of situations the empirical Bayes strategies are beneficial, since the resulting Bayesian (biased) parameter estimates often are better (in some senses) than their unbiased counterparts (Carlin and Louis, 2008). One such case is linear regression, where some common empirical Bayes strategies leads to shrinkage estimators highly related to ridge regression estimators. In Gangsei et al. (2016) it is shown how this 'empirical Bayes machinery' might be used in a situation with a bivariate response variable with missing data in order to increase prediction precision.

The denominator in the iconic Bayes theorem is known as the model evidence or the marginal likelihood. It might be viewed as a normalizing constant that ensures that the posterior distribution integrates to one over the model parameters. An alternative interpretation for the model evidence is 'likelihood of data conditional on model and hyperparameters'. The well–known Bayes factor for comparing two models is simply the ratio of their model evidences. Thus, when the model evidence has an analytic solution, as is the case for the present study, the natural (Bayesian) model selection method is to use model evidence (Kass and Raftery, 1995).

Posterior means are used as point-estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and 95% Credibility intervals are calculated by Monte Carlo sampling based on 10^4 simulations.

2.3 Model comparison

A total of 177 different combinations of predictor variables, denoted 'models' in the following paragraphs, were examined. These models included all 127 ($= 2^7 - 1$) possible combinations involving at least one of the predictor variables. Furthermore, 50 models involving interaction terms including the four continuous primary variables Fat, Meat, Totif and Weight as extra predictors were tested. These models were screened using BIC (Schwarz, 1978) as selection criteria via the R-package 'bestglm' (McLeod and Xu, 2011) for models using LMP–CT as

a single response variable. The final model was selected within these 50 screened models by examination of model evidence (Kass and Raftery, 1995) and 'RMSEP' (root–mean–squared–error of prediction).

Different models, and the difference between using the empirical Bayes method and ordinary least squares (OLS) based on the full observations, were compared by computing RMSEP via leave-one-out cross-validation. If not stated otherwise RMSEP was based on prediction results for manual dissection (LMP–MD).

To investigate the gain of choosing the empirical Bayes method utilizing the extra information from CT, over OLS, in situations with different combinations of sample sizes for LMP–MD and LMP–CT, a resampling study was conducted on the model with the favoured combination of predictor variables; Fat, Meat, Totif and Year. RMSEP was calculated based on regression parameter estimates from OLS and the empirical Bayes method using sets of data where $n_2 = 10, 15, 20, 30, \ldots, 80$ observations of LMP–MD were assumed known. The subsets were sampled randomly from the real data on two restrictions; the sample should contain at least 2 carcasses from each sex, and the overall average Fat value was to fall inside the range of Fat–values in the sample.

For every subsample of LMP–MD the corresponding set of LMP–CT values were assumed known. In addition $n_1 - n_2 = 0, 10, 15, 25, 50, 100, 200, 300$ extra observations of LMP–CT were assumed to be known, by random sampling from the real data. For all combinations of n_2 and n_1 RMSEP was calculated using OLS, and the empirical Bayes method. The calculations were based on a test set comprised by the $86 - n_2$ ($n_2 = 86$ corresponds to the full set of data) observations of LMP–MD discarded from the sub-sample used for parameter estimation. The process was repeated 500 times for every value of n_2, except for $n_2 = 80$ where the process was repeated 1000 times due to the small test set size for this value of n_2.

3 Results

3.1 Prediction precision and model selection

As part of the preliminary work for this paper, the data were also analysed using two alternative methods; two-stage least squares regression (2SLS) (Wooldridge, 2012, chap.15) and as a random effect model. To use the random effects model we had to assume that

$\sigma_{12} > 0$ and that $\sigma_{11} = \sigma_{22}$. Both these methods utilize the full set of data. The unreported results from these methods differed negligibly from the results obtained by the empirical Bayes method.

The model referred to as 'the favoured model', is the model where HGP7 variables and Gender were used as predictor variables. Among the models not including interaction–terms this model had the highest model evidence for hyperparameters; $\gamma_1 = 1.90$, $\gamma_2 = 1.22$, $\alpha_1 = 2.06$ and $\alpha_2 = 2.65$. The hyperparameters might be given an interpretation as prior population size, see (Gangsei et al., 2016) for detailed interpretation. Table 1 shows estimates for regression parameters using the favoured model. When interaction terms are not included the favoured model minimizes RMSEP when using both OLS and the empirical Bayes method. This model, and the model where Gender was omitted, are marked with arrows in Fig. 1.

Models including Fat as a predictor variable show a clear pattern of having much higher model evidence and lower values for RMSEP compared to models where Fat is excluded, c.f. Fig. 1. Inclusion of Meat and Totif as predictor variables unambiguously improves the model additionally. Among the factor variables Gender seems to be the only variable improving the predictive precision for LMP–MD as it is the only factor variable in the favoured model using model evidence as selection criteria. Furthermore, Gender is close to significant using OLS (p=0.053). RMSEP decreased from 1.69 to 1.67 (OLS) and 1.67 to 1.63 (empirical Bayes) when Gender was added as predictor variable.

Table 1 approx here.

Figure 1. approx here

When interaction terms were included as predictor variables, the model yielded no substantial improvement for prediction precision. A total of 13 models including interaction-terms had marginally larger model evidence and 12 models had smaller RMSEP than the favored model. The smallest RMSEP using interactions was 1.62 compared to 1.63 for the favoured model.

LMP–MD and LMP–CT were highly correlated, $\rho = .968$ (n=86). The effect of using the empirical Bayes method might be viewed as a way of borrowing strength from CT-data for estimating the parameters associated with manual dissection. The high correlation between LMP–MD and LMP–CT is exploited and increases the effective test–sample size. Thereby the variance of prediction parameter estimates are reduced and prediction precision

is increased.

Table 1 shows 95% Credible Intervals for parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ for the favoured model. Notice that the posterior distributions for the different elements of both $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in general depend on each other. This is analogous to the situation in OLS where $cov\,\hat{\boldsymbol{\beta}} = \sigma^2 \left(\boldsymbol{X}^t \boldsymbol{X}\right)^{-1}$.

The parameter $\sigma_{22} = \sigma_{MD}$ is naturally interpreted as the expected variance for the prediction error for LMP–MD conditional on known $\boldsymbol{\beta}_{MD}$ and known values for the predictor variables ($\boldsymbol{X}$). Consequently, for the favoured model, the lower limit for expected RMSEP was estimated to 1.63, i.e. the square root of posterior mean for σ_{22} (see Table 1). This is equal to the observed RMSEP. However, do note that the empirical Bayes method is a method specially designed to reduce prediction errors. Thus the parameter estimates, both for $\boldsymbol{\beta}$ and for $\boldsymbol{\Sigma}$ are (very mildly) biased.

The conditional variance, i.e. conditional on known regression parameter and predictor variables, of LMP–MD (σ_{22}) was smaller than the conditional variance of LMP–CT (σ_{11}) as a 95% Credible Interval for the fraction σ_{22}/σ_{11} was given by (0.57, 0.79), with a posterior median at 0.67. This corresponds to higher observed RMSEP-values for LMP–CT, typically around 1.9 for the best models.

3.2 Sufficient sample size

Figure 2 approx here.

Figure 2 shows the relationship between RMSEP and sample sizes for LMP–MD and LMP–CT, using the two methods OLS and empirical Bayes. The figure shows that when there were no additional observations for LMP–CT, OLS generated regression parameter estimates just as good as, or even better than, the empirical Bayes method. However, when additional observations of LMP–CT were present the empirical Bayes method generated regression parameter estimates yielding increased prediction precision compared with OLS–estimates.

4 Discussion

As the data structure with more observations for LMP–CT than LMP–MD is likely to be a challenge in numerous countries inside EU it might be advantageous to include methods

for analysing such data in 'Statistical handbook for assessing pig classification methods' (Causeur et al., 2003). A prerequisite for such inclusion is that there exist suitable software, like R–packages. Unfortunately a R–package for the empirical Bayes method applied in this study is not yet available. For 2SLS and random effects models software exists (Henningsen and Hamann, 2007; Pinheiro et al., 2015; Bates et al., 2015). Diggle et al. (1994) outlined a method for analysing the model in (1) by frequentist principles.

The difference between the regression parameter estimates using the empirical Bayes method and OLS was close to negligible. The favoured model, i.e. the model using HGP7-variables and Gender as predictor variables stood out as the better model, independently of method used for model selection (model evidence or RMSEP). The model omitting Gender as a predictor variable performed almost as well as the favoured model, and has the advantage of simpler data sampling since gender doesn't have to be registered.

The signs of the regression parameter estimates are in line with prior expectations. A thick Fat layer decreases predicted LMP, and oppositely a thick Meat layer increases expected LMP. The inner layer of fat (Totif) is the least important predictor variable from the HGP7. Its effect in the regression equation is likely to be severely affected by the other more dominant predictor variables. Finally the results show that conditional to similar measurements for HGP7-variables, females were expected to have a higher LMP than castrates.

When Year was included in the analysis the motive was not increased prediction precision as we have no observations for years 2014, 2015 etc. in the training data. Thus it is not possible to include Year as a predictor in the future. The motive for including Year in the analysis was to see if any effect was present, effects that are not a result of the Year itself. Such effects might have numerous explanations like differences in the manual– or CT–dissection (for instance different butcher teams), development of the body composition in pigs (for instance due to breeding/genetic effects) etc. As the effect of Year in the present study is close to negligible it indicates that the validity of the prediction equation is applicable over time. This interpretation is subject to considerable uncertainty as only two different years, 2008 and 2013, is involved in the study.

Omitting Gender as a predictor variable is in line with the principles used in the Netherlands, where the effect of gender on the prediction equation using HGP7 has been evaluated in detail by Engel et al. (2006, 2012). Engel et al. (2012) was a study aiming at finding robust methods for handling different proportions of females, castrates and males in the

pig population. They found a significant effect of Gender, but ended up using a prediction equations for HGP7 where Fat and Meat were the two only predictor variables, data for the inner fat layer (Totif) was not used in their analysis. Engel et al. (2006, 2012) and Font i Furnols and Gispert (2009) reported RMSEP at 2.24, 2.10 and 1.8 % units respectively, a little larger than the RMSEP at 1.63, reported for Norway in the present study. In Font i Furnols and Gispert (2009) Fat-O-Meater, an optical probe similar to HGP7, was used.

The results from this study show a significant effect of Gender as a predictor variable for LMP–MD. Due to considerable extra costs if Gender is to be sampled for every carcass, models omitting Gender might be preferable. Since the bulk of pig farmers deliver a close to equal proportion of females and castrates to slaughterhouses, omitting Gender as a predictor variable will have minor effect on the total cash settlement between farmer and slaughterhouse. However, the omission will lead to a bias where LMP, on average, will be underestimated for females and overestimated for castrates.

The effect of Gender indicates differences between females and castrates regarding the meat and fat distribution in the carcass. If these differences affect the profitability of the possible processing methods, it might be profitable for slaughterhouses to register Gender. Further analysis of such effects falls outside the scope of this study.

Since the distribution of fat and meat in the carcasses depends to some extent on the gender, it is very likely that such dependencies might also occur between breeds, even if the effect was ignorable for the three breeds evaluated in the present study.

An eventual effect of Breed, unlike Gender, would introduce a bias providing a systematic effect, positive or negative, for different farmers and cooperatives using different breeds. Since LMP differs between breeds, see appendix A.1, there is a consistent demand from different cooperatives and farmers that the effect of Breed is to be tested and accounted for in the prediction equation. Consequently, the ignorable effect of Breed demonstrated for the present study does not rule out testing for this effect in forthcoming updates of the prediction equation for LMP.

Furthermore, there seems to be very limited gain in including interaction terms. Such inclusions make the models more complex, and thereby increase the possibility for substantially biased prediction of carcasses having anomalous on–line measurement values. The non usefulness of second order interaction terms strongly suggests that including higher order interaction terms or quadratic terms would not be beneficial.

The estimated values for the different elements of $\boldsymbol{\beta}$ were very similar for the estimates regarding LMP–MD and LMP–CT, with exception for the intercept term, which was larger for LMP–CT. The natural interpretation is that this difference reflects the difference of 6.3 percent units between average observed LMP–CT and observed LMP–MD.

Statistically, the expected value for RMSEP is the square root of the sum of 'squared bias' and 'error variance' for the predicted values. For any given model, the only way to reduce the expected value for RMSEP is to reduce 'squared bias' as the (model specific) 'error variance' is assumed to be fixed.

In the present case, with 86 observations of LMP–MD, and a fairly simple model including only four predictor variables, the bulk of the prediction errors are due to the modelled error variance, and not a result of biased regression parameters. Consequently, the empirical Bayes method did not substantially improve prediction precision for LMP–MD in terms of RMSEP.

The usefulness of applying the empirical Bayes method, or other methods utilizing the information from a surrogate variable, i.e. LMP–CT, depends heavily on the sample sizes of observed LMP–MD and LMP–CT, and the covariance-matrix for the error terms ($\boldsymbol{\Sigma}$). The relationship between sample sizes of manually dissected carcasses, CT scanned carcasses and RMSEP shown in Fig. 2 might be used to optimize sample sizes if the cost of sampling and the gain of reduced expected prediction error is known. The empirical Bayes method might be applied to heavily reduced sets of data without appreciable loss of prediction precision.

In situations where several breeds are present methods utilizing CT–data will be especially useful. Then the training set might be composed of a limited number of manually dissected carcasses, and a larger number of CT scanned carcasses containing a sufficient number of carcasses from all breeds.

Prior to looking at the results, we had expected higher precision for predicted LMP–CT values compared with LMP–MD values, due to an assumption that CT is a more precise method for predicting LMP than manual dissection. Evaluation of the manual dissection method shows a generally high accuracy and reliability for the estimated LMP–MD (Nissen et al., 2006), but also revealed some problematic issues, for instance a significant effect of butcher for the estimated LMP–MD.

The results in this study show the opposite, that on–line measurements using HGP7 tend to predict LMP–MD more precisely than LMP–CT. This is seen by evaluating the variances for the errors, and by comparing RMSEPs for LMP–MD and LMP–CT. The pattern is

evident even if the sample is restricted to the carcasses where both LMP–MD and LMP–CT are observed. Thus, the non random sampling of carcasses for manual dissection does not explain the observation. However, the difference might, at least partly, be explained by the fact that the manual dissection method is a partial dissection method (Walstra and Merkus, 1995), whereas CT performs a total dissection. Thus the measurements obtained by the HGP7 on the back of the carcass might be better correlated with the partial manual dissection than the full CT–dissection.

The estimate for $\sigma_{22} = \sigma_{MD}$, and the estimated RMSEP values indicates that the EU specification of an RMSEP at maximum 2,5 is easily fulfilled using HGP7 on Norwegian carcasses.

5 Conclusion

A model using four predictor variables, Fat, Meat, Totif (from the optical probe – HGP7) and Gender, is simple and provides a high prediction precision well inside EU standards. The three variables Year, Breed and Weight seems to be of minor or no importance for LMP–MD prediction when combined with the HGP7 variables and Gender for the data set used in the present study. This result should not be generalized without reservation. Inclusion of second order interaction terms does not improve prediction precision substantially.

For analytical simplicity OLS-regression using input from the optical probe and Gender is the better method and model. However, for practical simplicity, Gender might be omitted without severe loss of prediction precision.

The drawback of OLS regression is its inability to utilize extra information from CT scanned carcasses. In order to limit the needed sample size of manually dissected carcasses regression methods utilizing information from CT might be applied to a training sample with a low number of manually dissected carcasses, but a sufficiently large number of CT scanned carcasses, yielding no loss or minor loss of prediction precision.

6 Acknowledgments

Training schools and seminars arranged by the EU-founded COST Action FAIM have provided valuable knowledge for the present study. Trygve Almøy and Frøydis Bjerke have

contributed to this article by useful comments to the manuscript. Finally, thanks are due to all personnel at the slaughterhouses, Animalia and DMRI, for their patience and effort that made the data sampling possible.

References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed–Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Carlin, B. P. and Louis, T. A. (2008). *Bayesian Methods for Data Analysis*. CRC Press.

Causeur, D., Daumas, G., Dhorne, T., Engel, B., Font i Furnols, M., and S, H. (2003). Statistical handbook for assessing pig classification methods: Recommendations from the "EUPIGCLASS" project group.

Commission of the European Communities (2006). Commission Regulation (EC) No 1197/2006 of 7 august 2006 amending Regulation (EEC) No 2967/85 laying down detailed rules for the application of the Community scale for grading pig carcases. *Official Journal of the European Union*, L 217(49):6–7.

Commission of the European Communities (2008). Commission Regulation (EC) No 1249/2008 of 10 December 2008 laying down detailed rules on the implementation of the Community scales for the classification of beef, pig and sheep carcases and the reporting of prices thereof. *Official Journal of the European Communities*, L 337:3–30.

Daumas, G. and Monziols, M. (2011). Comparison between computed tomography and dissection for calibrating pig classification methods. In *57th International Congress of Meat Science and Technology (ICoMST 2011)*, volume 1, pages 296–299. Ghent-Belgium, 7–12 August 2011.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of longitudinal data*. Oxford University Press, Oxford,UK.

Engel, B., Lambooij, E., Buist, W., Reimert, H., and Mateman, G. (2006). 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. *Animal Science*, 82(6):919–928.

Engel, B., Lambooij, E., Buist, W., and Vereijken, P. (2012). Lean meat prediction with HGP, CGM and CSB-Image-Meater, with prediction accuracy evaluated for different proportions of gilts, boars and castrated boars in the pig population. *Meat science*, 90(2):338–344.

Font i Furnols, M. and Gispert, M. (2009). Comparison of different devices for predicting the lean meat percentage of pig carcasses. *Meat science*, 83(3):443–446.

Gangsei, L. E., Almø y, T., and Sæ bø, S. (2016). Linear regression with bivariate response variable containing missing data. an empirical Bayes strategy to increase prediction precision. Submitted manuscript to Communications in Statistics – Simulation and Computation.

Henningsen, A. and Hamann, J. D. (2007). systemfit: A Package for Estimating Systems of Simultaneous Equations in R. *Journal of Statistical Software*, 23(4):1–40.

Judas, M., Höereth, R., and Branscheid, W. (2007). Computed tomography as a method to analyse the tissue composition of pig carcases. *Fleischwirtschaft International*, 1:56–59.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.*, 90(430):773–795.

McLeod, A. and Xu, C. (2011). *bestglm: Best Subset GLM*. R package version 0.33.

Nissen, P. M., Busk, H., Oksama, M., Seynaeve, M., Gispert, M., Walstra, P., Hansson, I., and Olsen, E. (2006). The estimated accuracy of the EU reference dissection method for pig carcass classification. *Meat science*, 73(1):22–28.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-121.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The annals of statistics*, 6(2):461–464.

Vester-Christensen, M., Erbou, S. G., Hansen, M. F., Olsen, E. V., Christensen, L. B., Hviid, M., Ersbøll, B. K., and Larsen, R. (2009). Virtual dissection of pig carcasses. *Meat science*, 81(4):699–704.

Walstra, P. and Merkus, G. (1995). *Procedure for the assessment of lean meat percentage as a consequence of the new EU reference dissection method in pig carcass classification.* DLO Research Institute of Animal Science and Health (ID-DLO), Zeist, The Netherlands.

Wooldridge, J. M. (2012). *Introductory econometrics: a modern approach.* Cengage Learning.

A Appendix

A.1 Summary statistics for response and predictor variables

Table A.1 approx here

Table 1: 95% Credible Intervals, and posterior median and mean values, for the elements of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ using the favoured model.

$\boldsymbol{\beta}_{CT}$	2.5%	50%	97.5%	Mean
Intercept	66.06	66.31	66.56	66.31
Fat	-1.18	-1.11	-1.04	-1.11
Meat	0.07	0.10	0.13	0.10
Totif	0.06	0.11	0.16	0.11
Gender	0.19	0.56	0.94	0.56
$\boldsymbol{\beta}_{MD}$	2.5%	50%	97.5%	Mean
Intercept	60.23	60.51	60.77	60.51
Fat	-1.10	-1.03	-0.95	-1.03
Meat	0.07	0.10	0.13	0.10
Totif	0.08	0.13	0.18	0.13
Gender	0.31	0.71	1.12	0.71
$\boldsymbol{\Sigma}$	2.5%	50%	97.5%	Mean
σ_{11} (CT)	3.17	3.60	4.09	3.61
σ_{12}	2.01	2.41	2.93	2.42
σ_{22} (MD)	2.28	2.67	3.12	2.68

Table A1: Mean and standard deviation, in parenthesis, for LMP–MD (%), LMP–CT (%) and the 4 continuous predictor variables (Fat (mm), Meat (mm), Totif (mm) and Weight (kg)). Values distributed over the two levels for Year, i.e. 2008 and 2013, and the three levels for Breed, i.e. Hampshire, Norhybrid and Duroc. The numbers n_{CT} and n_{MD} shows sample sizes for CT scanned and manual dissected carcasses respectively.

2008	**Hamp.**	**Norhyb.**	**Duroc**	**All**
n_{CT}	37	75	117	229
n_{MD}	10	20	36	66
LMP–CT	66.6 (3.1)	66.5 (4.1)	66.4 (3.3)	66.4 (3.5)
LMP–MD	60.5 (3.6)	60.6 (4.1)	60.1 (2.4)	60.3 (3.1)
Fat	11.9 (2.1)	12.5 (3)	12.4 (2.3)	12.3 (2.5)
Meat	57.9 (6.8)	54.9 (5.5)	56.1 (6)	56 (6)
Totif	12.7 (4.9)	12.6 (3)	10.6 (4.3)	11.6 (4.2)
Weight	81.1 (6)	82.1 (6.8)	78.4 (5.8)	80.1 (6.4)

2013	**Hamp.**	**Norhyb.**	**Duroc**	**All**
n_{CT}	32	78	126	236
n_{MD}	10	0	10	20
LMP–CT	66.6 (2.6)	67.5 (4)	66.3 (3.7)	66.7 (3.7)
LMP–MD	60.6 (2.8)	-	59.8 (3.5)	60.2 (3.1)
Fat	11.9 (2.2)	11.3 (2.8)	12.2 (2.7)	11.8 (2.7)
Meat	59.6 (6.2)	55.7 (6.1)	53.3 (7.1)	55 (7)
Totif	10.8 (3.7)	13 (4.2)	10.6 (3.9)	11.4 (4.1)
Weight	80.8 (6.6)	79.2 (8.6)	76.1 (7.6)	77.7 (8)

Both years	**Hamp.**	**Norhyb.**	**Duroc**	**All**
n_{CT}	69	153	243	465
n_{MD}	20	20	46	86
LMP–CT	66.6 (2.9)	67 (4.1)	66.3 (3.5)	66.6 (3.6)
LMP–MD	60.5 (3.1)	60.6 (4.1)	60.1 (2.6)	60.3 (3.1)
Fat	11.9 (2.2)	11.9 (3)	12.3 (2.5)	12.1 (2.7)
Meat	58.7 (6.5)	55.3 (5.8)	54.7 (6.7)	55.5 (6.5)
Totif	11.8 (4.5)	12.8 (3.7)	10.6 (4.1)	11.5 (4.2)
Weight	80.9 (6.2)	80.6 (7.9)	77.2 (6.9)	78.9 (7.3)

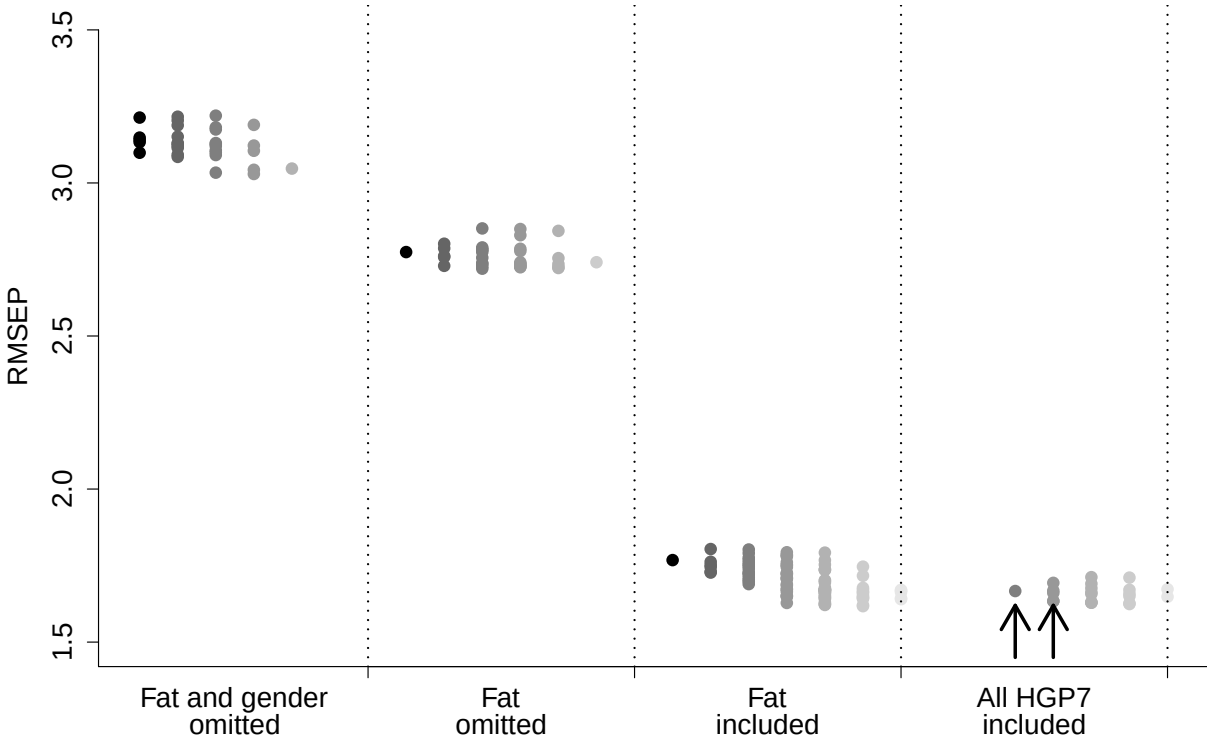

Figure 1: Relationship between number of predictor variables and RMSEP, using the empirical Bayes method, for all 127 possible model-combinations including of the primary predictor variables, and the 50 models including interaction-terms. Models omitting both Fat and Gender as predictor variable are shown in the leftmost group. Models where Gender, but not Fat, was included are shown in the second leftmost group. Models where Fat was included, but where Meat and/or Totif were omitted are shown in the middle group. The second rightmost group represents models where all HGP7-variables (Fat, Meat and Totif) were included, and finally the rightmost group represents the 50 models where interaction terms were included. Black colour shows models with one predictor variable, then the gradual transition to the lightest grey represents 2,..., up to a maximum of 7 predictor variables. The two black arrows points at the favoured model (right arrow) and the model where Gender is omitted (left arrow).

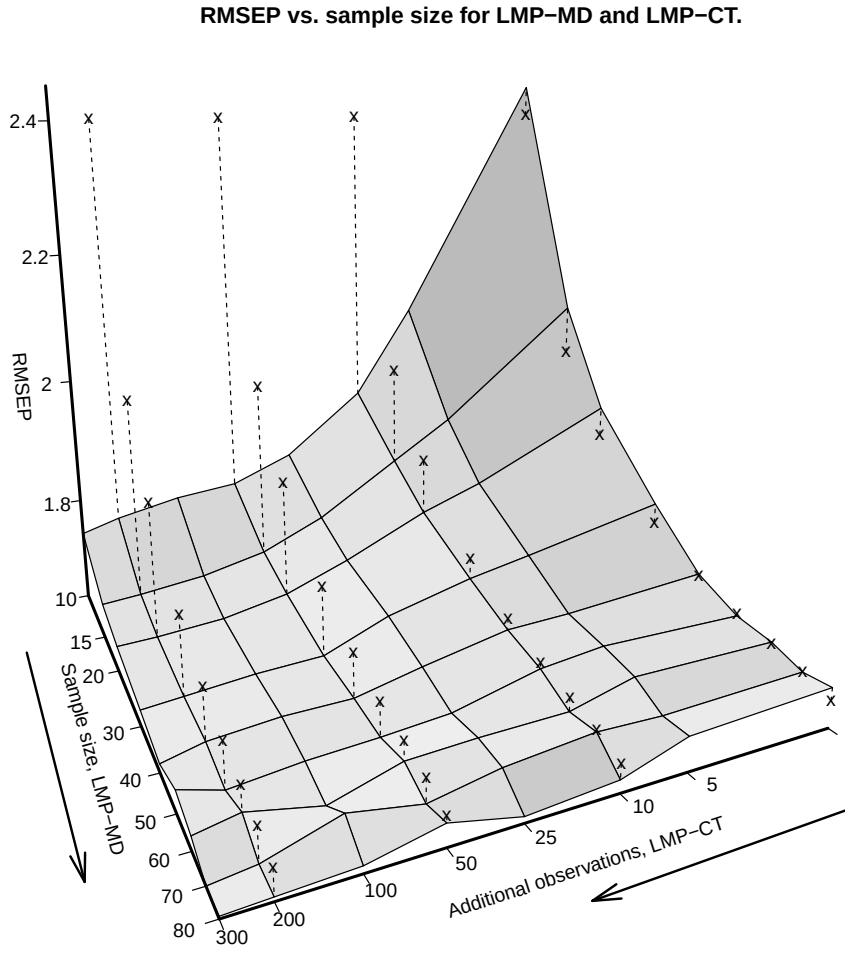

Figure 2: RMSEP for LMP–MD based on samples with varying number of observations for LMP–MD and additional observations of LMP–CT. Each sample size combination for LMP–MD and LMP–CT represent a 'line-crossing' in the plane, and is represented with the average RMSEP value for 500 (1000 for LMP–MD size at 80) different random samples of LMP–MD and LMP–CT using the empirical Bayes model. Points represented by 'x'-s show corresponding RMSEP–s using OLS. The dotted lines represent the difference (i.e. the gain) of using the empirical Bayes method instead of OLS for each combination of sample-sizes for LMP–MD and LMP–CT.