Contemporary Groups Are Always Random

L. R. Schaeffer

February 18, 2009

1 INTRODUCTION

Since 1970 when Dr. Henderson introduced the Northeast AI Sire Comparison model and methods to the world, herd-year-season effects (contemporary groups (CG) in that model) have been treated as fixed effects. This was based on Dr. Henderson's selection bias theory in which the bias being considered was that the true sire genetic values were associated with the true herd-year-season (HYS) effects. That is, herds with better management effects on their cows tended to use the genetically higher valued sires on their cows to produce the first lactation daughters in a herd. Henderson naturally thought of herd-year-season effects as random (as discussed in a video from Kyoto, Japan in 1985), but his selection theory told him that either HYS or sire effects needed to be treated as fixed effects to eliminate the bias from the model. Because sire genetic values were more important at the time, he chose to make HYS fixed rather than sire effects. Subsequently, every genetic evaluation system has used fixed CG, even though it does not make sense to do so in all situations.

There was never a study to measure the amount of bias that existed if HYS effects were random, nor any quantification of the degree of association between true (unknown) HYS effect and true (unknown) sire effects. This was assumed to be a known problem in some (not all) herds. Another assumption was the Dr. Henderson's selection theory was correct. This was logical because Dr. Henderson was invariably correct about everything. His selection theory is now considered incorrect and has been seriously questioned by several researchers (including Robin Thompson, Richard Quaas, and Daniel Gianola, if not others). Still the tradition persists and there has been no mad rush to make CG random effects again.

Even if Henderson's selection theory was correct, is it logical to presume in all species and for all traits that the CG should be treated as fixed. That is, is there actually any bias present for that trait, in that species. For example, sheep breeding in Canada is such a minor agricultural enterprise with little information about rams and ewes prior to matings. Often a breeder has limited numbers of rams to choose as mates for their ewes. An association between ram genetic values and flock CG effects is difficult to visualize across flocks. Another example is health traits in dairy cattle. There has been virtually no information about the health of daughters of bulls, or the health level of herds to contemplate an association between herd CG effects for health and sire true genetic values for health.

1.1 Questions

- How should the sire by CG associations be quantified?
- What is the amount of bias when CG are random effects?
- Are biases completely removed by treating CG as fixed effects?
- How much do the standard errors of prediction increase on sire ETA when CG are treated as fixed effects?
- Are there other alternatives to remove the bias?

1.2 Definition of Contemporary Groups

Contemporary groups (CG) are defined and formed by researchers to identify a group of animals that are roughly of the same age and sex, and that have undergone the same management and environmental conditions during a particular phase of their life (Van Vleck, 1987). CG have typically been formed on the basis of herd (H), year of calving (Y), and season of calving (S) or HYS effects for genetic analyses of 305-d lactation production by an animal model.

The CG effect can be viewed as an H(erd) effect nested within YS(Year-season) subclasses. The YS subclasses are the main effects that monitor time trends over years and seasons, and herd effects are nested within YS subclasses because the herd management effect is continuously changing with time. Managers do not apply the exact same feeding system over several years. The feeds that were utilized would not be identical in quality over the years. Quota decisions, water quality, diseases, milking machines, housing systems and other variables are constantly changing as well as the cows themselves. The only constant variables of a herd are the owner, the herd identification, and the herd location (and these could change too).

The YS effects could be different for various regions(R) of a country, and therefore, a more reasonable main effect would be a RYS subclass effect, with herds nested within these subclasses. A CG effect, in this paper therefore, will be a RYS:H where (:H) means H nested within the larger effect.

2 Modelling CG

From an estimability perspective, a model with fixed RYS:H effects is identical to a model with fixed R, Y, S plus all two-way, and three-way interactions, with H nested within each of these. The estimate of a particular RYS:H effect, therefore, contains the true effects of each of the main effects, two-way and three-way interaction effects.

$$y_{ijkl} = R_i + Y_j + S_k + RY_{ij} + RS_{ik} + YS_{jk} + RYS_{ijk} + RYS : H_{ijkl} + \text{other stuff.}$$

The above model is equivalent to

$$y_{ijkl} = RYS : H_{ijkl} + \text{other stuff.}$$

When CG are random, then RYS:H is random, and the model should be modified to include fixed RYS subclass effects because random RYS:H does not account for the RYS subclass effects. The model should be

$$y_{ijkl} = (RYS)_{ijk} + (RYSP : H)_{ijkl} + \cdots,$$

The fixed $(RYS)_{ijk}$ effects include R_i, Y_j, S_k, RY_{ij}, RS_{ik}, and YS_{jk} effects, which are non-estimable.

This has been a common oversight in simulation studies that have compared fixed versus random contemporary group effects in sire models, i.e. the (RYS) effects are usually not included in the model when RYS:H effects are random, and consequently, biases in estimated breeding values become very large.

3 Henderson's Theory

The proof that treating YS:H effects as fixed provides unbiased predictions of sire transmitting abilities in a sire model with unrelated sires was given by Henderson (1973, 1975) based upon Pearson's (1903) selection model. Thompson (1979) questioned the validity of conditioning expectations on the selection differentials, and was confused by the matrix $\mathbf{L}$, which Henderson (1975) used to describe the selection process, because it was a random variable and not fixed.

Gianola et al. (1988) went further by arguing against the idea of conceptual repeated sampling which Henderson (1975) found necessary to invoke. Gianola et al. (1988) pointed out that when Henderson (1973) required predictors of sire transmitting abilities to be unbiased by treating YS:H effects as fixed, that this did not necessarily maximize expected genetic response from selection on sire estimated transmitting abilities. Bayesian methods were recommended to perhaps get around these problems.

The conclusion from these two papers was that the procedures given by Henderson (1975) may not be appropriate. The theoretical basis for treating YS:H effects as fixed is therefore, put in doubt.

4 Simulation Studies

Two simulation studies by Ugarte et al. (1992) and by Visscher and Goddard (1993) compared fixed versus random CG effects on the bias and accuracy of sire estimated transmitting abilities. Both studies used a simple sire model with sires assumed to be unrelated. Neither study incorporated time trends or selection of cows and mates over years in their simulations. Therefore, there were no year of calving, season of calving, or herd effects involved in the simulations, and therefore, no need to account for time trends (i.e. YS effects) in their data. The results of Ugarte et al. (1992) basically supported the contentions of Henderson (1973), but concluded that benefits could arise from using random CG effects in certain situations depending on CG subclass size and parameter values. Visscher and Goddard (1993) took an analytical approach to show that the results of Ugarte et al. (1992) were not general, and how sire estimated transmitting abilities could be negatively correlated to true transmitting abilities when CG effects were random depending on the nature of the sire by CG associations. To date, there has been no study of fixed or random CG in animal models in which time trends and cow and sire selection have been incorporated.

5 Herd by Sire Associations

Herds vary in the wealth and business acumen of the owners. Thus, wealthy owners can afford to pay for superior sires while less fortunate owners may only afford to pay for young sires. Thus, the association between herds and sires is most likely between herd wealth and sire superiority (for traits known only to the herd owner). There is not really an association between the true unknown RYS:H effect and sire superiority unless there is a high correlation between herd wealth and true RYS:H effects. When the term 'better herds' is used, this usually implies the more wealthy herds, rather than any reflection on the true RYS:H effect (which could actually be negative even in a well-managed and wealthy herd, or positive in a poorly-managed herd). Thus, there is confusion generated by terminology. The correlation between herd wealth (or management ability) and true RYS:H environmental effects is really unknown and could be close to zero.

Henderson (1988) presented a method for measuring bias, but this method requires the inverse elements of the mixed model equations, which are invariably impossible to obtain from animal models.

Sires are generally highly selected and perhaps sire effects should be treated as fixed in an animal model rather than contemporary groups (Meyer, 1982; Van Vleck, 1985). Not all of biases can be removed statistically. Preferential treatment of daughters of particular sires is difficult to define and measure (Tierney and Schaeffer, 1994).

The importance of a sire by herd (wealth or management) association is deemed significant by most animal breeders, but there are no estimates of the magnitude of the association or of the bias that it might cause in genetic evaluations.

6 My Advice

As long as you are not working with Holstein production data (and even if you are),

- Always start with contemporary groups as a random factor in the model, and also include a fixed factor of time in the model, such as year-seasons, or year-months. This time factor is crucial to avoiding an overestimation of genetic trends.
- Test for an association of sires with contemporary groups before making the contemporary group factor a fixed factor. Correlate sire progeny averages with contemporary group averages (for the trait of interest), and plot sire averages against contemporary group averages for the contemporary groups in which sires appear.
- If an association exists, think about making sires a fixed factor in the model rather than contemporary groups.
- Another test might be to compute the variability of residuals within contemporary groups (as random effects) and within sire progeny groups, and plot against each other. A correlation of zero indicates no association.

The amount of actual bias is probably very small, and the correlation of estimated breeding values with true breeding values should be higher than when contemporary groups are treated as fixed.

Conceptually, contemporary group effects are truly random. No one can predict what the effect is going to be before it happens. On the other hand, the sires that are used by breeders are probably more "known" because either proofs or parent averages or genomic breeding values are available. Producers can make a deliberate decision about using a sire, and therefore, sire effects are not random. They should be fixed. The progeny that are produced should be a random sample of progeny of that sire (unless they are also genotyped at birth and selected on that basis to make a record in the herd).

7 REFERENCES

Gianola, D., S. Im, and R. L. Fernando. 1988. Prediction of breeding values under Henderson's selection model: A revisitation. J. Dairy Sci. 71:2790-2798.

Henderson, C. R. 1973. Sire evaluation and genetic trends. Proc. Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush. ASAS, ADSA, PSA, pp 10-41.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-448.

Henderson, C. R. 1988. Simple method to compute biases and mean squared errors of linear estimators and predictors in a selection model assuming multivariate normality. J. Dairy Sci. 71:3135- 3142.

Meyer, K. 1982. Estimation of genetic parameters for later lactations in dairy cattle. Second WCGALP, Madrid. 7:256-260.

Pearson, K. 1903. Mathematical contributions to the theory of evolution. XI On the influence of natural selection on the variability and correlation of organs. Philosophical Transactions of the Royal Society of London, A, 200: 1.

Thompson, R. 1979. Sire evaluation. Biometrics 35:339-353.

Tierney, J. S. and L. R. Schaeffer. 1994. Inclusion of semen price of the sire in an animal model to account for preferential treatment. J. Dairy Sci. 77:576-582.

Ugarte, E., R. Alenda, and M. J. Carabano. 1992. Fixed or random contemporary groups in genetic evaluations. J. Dairy Sci. 75: 269-278.

Van Vleck, L. D. 1985. Including records of daughters of selected bulls in estimation of sire component of variance. J. Dairy Sci. 68:2396-2402.

Van Vleck, L. D. 1987. Contemporary groups for genetic evaluations. J. Dairy Sci. 70:2456-2464.

Visscher, P. M. and M. E. Goddard. 1993. Fixed and random contemporary groups. J. Dairy Sci. 76:1444-1454.