ESTIMATION OF VARIANCE COMPONENTS FOR THE U.S.CONSUMER PRICE INDEX

Owen J. Shoemaker and William H. Johnson, U.S. Bureau of Labor Statistics
2 Mass Ave., NE, Room 3655, Washington, DC 20212

Key Words: Weighted REMLs, weighted least squares, sample design, PSU, ANOVA

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

For every new sample for the commodities and services (C&S) component of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics produces a C&S sample design in which the outlets and items are allocated in an optimal fashion. This item-outlet optimization C&S sample design requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. A fourth component of variance is the error term. The total variance of these unit components of variance, divided by their respective number of PSUs, item-strata, outlets and quotes, is then minimized by the optimal number of respective outlets and item hits, as constrained by a cost function.

1. The Design

Commodities and Services (C&S) accounts for 72.5% of the CPI (as measured in expenditure shares), with Housing accounting for the remaining 27.5%.[1] The first stage of the overall design is the PSU (Primary Sampling Unit) sample selection. This stage is common to both Housing and C&S. The CPI is conducted in 87 PSUs in 83 geographic areas. (New York City has three PSUs, Los Angeles two and Baltimore-Washington two.) The 31 largest A-level PSUs are selected with certainty. The 56 smaller B- and C-level PSUs are then selected with probability proportional to size[2] (pps) within their respective regions: Northeast, Midwest, South or West. The Index itself is published at the AREA level, which includes all of the A-level PSUs. For the B- and C-level PSUs the Index is calculated and published at the regional level: X100, X200, X300 and X499 are the respective AREAs for the B's in the four regions, and D200, D300 and D400 the AREAs for the C's (no C-level city in the Northeast got selected in the last pick). Thus PSU is a random factor only in the AREAs X100, X200, X300, X499, D200, D300 and D400.

Two further independent sampling stages occur in C&S: outlet allocation and item allocation, within each PSU-Replicate combination. The outlet sample is based on the Telephone Point of Purchase Survey (TPOPS), conducted by the Bureau of Census for the BLS. The item sample is based on the Consumer Expenditure Survey (CE), also conducted by Census for BLS. Outlets are selected in 217 TPOPS categories using a systematic pps[3] sampling scheme. Items are selected in 13 major groups using a stratified systematic pps sampling scheme. The components of variance themselves are calculated at the AREA level (i.e., all the A's plus the 4 X's and the 3 D's as delineated above) by Major-Group. For example, a set of three components of variance (ITEM, OUTLET & Error) is calculated for APPAREL (MAJ_GRP 7) in Atlanta (AREA A319), and a set of four components of variance (PSU, ITEM, OUTLET & Error) is calculated for, say, MEDICAL in X100 (the B-level cities in the Northeast). The estimator is the average price change for each AREA – MAJ_GRP category. The total variance of each estimator is modeled as the sum of the four components (or three, in the certainty A-level PSUs):

$$\sigma^2_{j,k} = \sigma^2_{psu,j,k} + \sigma^2_{item,j,k} + \sigma^2_{outlet,j,k} + \sigma^2_{error,j,k},$$

for each major-group j and area k. We then assume that the variance of price change of an individual sampled unit or quote has the same structure:

$$\sigma^2_{unit,j,k} = \sigma^2_{unit,psu,j,k} + \sigma^2_{unit,item,j,k} + \sigma^2_{unit,outlet,j,k} + \sigma^2_{unit,error,j,k}, \text{ and}$$

$$\sigma^2_{j,k} = \sigma^2_{unit,psu,j,k} / N_p + \sigma^2_{unit,item,j,k} / N_h + \sigma^2_{unit,outlet,j,k} / N_o + \sigma^2_{unit,error,j,k} / N_e,$$

[1] Bureau of Labor Statistics, *CPI Detailed Report (Feb 1999),* p. 7.

[2] Size here equals population.

[3] For OUTLET and ITEM sample selection size equals expenditures.

where each N_i stands for an appropriate number of psu's, items, outlets or quotes.

The sampling variance of price change for the All US C&S Index is $\sigma^2_{TOTAL} = \sum_j \sum_k RI^2_{j,k} \cdot \sigma^2_{j,k}$. The $RI_{j,k}$'s are the relative importances for each AREA – MAJ_GRP combination, factoring in both relative expenditures and relative populations, with $\sum_j \sum_k RI_{j,k} = 1$. It is this σ^2_{TOTAL} that is then minimized in the optimization procedure, but it is the unit-level components of variance that will be calculated and analyzed, using weighted REMLs (Restricted Maximum Likelihood Estimation) as our chosen methodology.

2. The Weights

Before proceeding to the general linear model analyses, we would like to present the rationale for and construction of the weights that we used. In the CPI not all price changes are equal. An elaborate system of weights, essentially based on expenditures shares, is used at both the basic and aggregate levels of index calculation. At the basic item-stratum level, within each AREA, a group of prices-with-weights are combined into a sub-index (or, price relative), using either a Laspeyres or a Geomeans formula. These price relatives are then aggregrated, this time with separate aggregate weights, into various higher level indexes which then become the particulars of the CPI, including the All US–All Items Index itself.

The random variable of interest for calculating our components of variance is not, however, this price relative, but an individual price change. (No OUTLET variance component could be produced without price change observations at the quote level and, besides, a unit component of variance requires exactly that: a unit-level observation.) But in line with the basic price relatives and the higher-level aggregate indexes, the individual price changes and their concomitant variance components invite a weight structure as well. Happily we do have basic weights (FNLWs) attached to each individual price quote. If we can configure these weights with appropriate relative importances, we might be in business. And we can. We have cost-weights for every item-stratum in each AREA by MAJ_GRP category and so we can calculate a relative importance ($RI_{is,j,k}$) for every item-stratum in each AREA by MAJ_GRP category. The resultant quote-weight is simply: QWT = FNLW * RI. The weights, however, are not attached to the observations themselves but to the residual sum of squares: i.e., $\Sigma w(y-\hat{y})^2$. For an ANOVA-based analysis this means using weighted least squares; for REML estimation the squared residuals themselves are weighted.

One last crucial decision had to be made however: whether to utilize these QWT's as is or to rescale them in some way. Pfeffermann (et al), in a 1998 paper, "Weighting For Unequal Selection Probabilities In Multilevel Models",[4] offers the mean of the weights as a candidate rescaling factor. Thus, $W_{i,j,k} = QWT_{i,j,k} / \lambda$, using Pfeffermann's suggested $\lambda = \sum_i QWT_{i,j,k} / n_{j,k}$. Besides "retaining consistency" and probably reducing small sample bias (the main object of Pfeffermann's inquiries and recommendations), this particular mean-scaled weight structure leaves the linear model results themselves at the same unit level at which we want our components of variance. I.e., $\sum_i W_{i,j,k} = n_{j,k}$, just as the trace(I) = $n_{j,k}$ in any unweighted model. Moreover, this re-scaling clearly retains the full information content of the original weights. Finally, the use of these weights seems to have produced two additional side-benefits: (1) more stable components of variance across time periods, and (2) a final result, the σ^2_{TOTAL}, that compares favorably and properly in magnitude with an independent variance calculation for the All U.S.– All Items Index.[5]

3. The Model

The model we will analyze treats all three effects as random. The design is unbalanced. We let y_{ijkl} be the observed unit price change between time t and time t–6 for quote l within psu i, item j, and outlet k. Then,

(1) $y_{ijkl} = \mu + p_i + h_j + o_k + e_{ijkl}$,

where μ is a fixed effect

$p_i \sim N(0, \ ^2_{unit,\ psu})$

$h_j \sim N(0, \ ^2_{unit,\ item})$

$o_k \sim N(0, \ ^2_{unit,\ outlet})$

[4] D. Pfeffermann, et al (1998), *Weighting For Unequal Selection Probabilities In Multilevel Models*, Journal of the Royal Statistical Society, Ser B, Vol 60, pp 23-40

[5] Using a stratified random group methodology from a replicate structured index data base, the All-US–All-Items Index variance measure (standard error) for a 1995 6-month price change was ≈ 0.15 as compared to the design's final optimal SE of ≈ 0.08.

$$e_{ijkl} \sim N(0, \ ^2_{unit,\ error}),$$

with p, h, o and e all independent of each other.

We can then write (1) in vector notation as

$\mathbf{y} = \mathbf{J}\mu + \mathbf{Z}\gamma + \mathbf{e}$, where

$\mathbf{y}$ is an n x 1 vector of the y_{ijkl}'s,
$\mathbf{J}$ is an n x 1 vector of 1's.
$\mathbf{Z}$ is a known n x s design matrix,
where s = P+H+O
$\gamma = (p_1, \ldots, p_P, h_1, \ldots, h_H, o_1, \ldots, o_O)^T$,
an s x 1 vector of random effects.
$\mathbf{e}$ is an n x 1 vector of the e_{ijkl}'s,
P is the number of psu's sampled,
H is the number of items sampled,
O is the number of outlets sampled.

A special case of this model, when there is only one PSU in the model, means $p_i = 0$, and the model can be then written:

$$y_{ijkl} = \mu + h_j + o_k + e_{ijkl} .$$

This special case model applies to all the A-sized AREAs. The main model encompasses all the rest: i.e., the X- and D-sized AREAs.

4. Choosing a Methodology

Utilizing this basic linear model, we then have several choices of methodologies. First of all, we could compute the components of variance using an ANOVA-based structure (Type I Method). Type I variance components use a sequential sum of squares methodology wherein the sum of the sum of squares for each effect (plus error) adds up to the total sum of squares of the model exactly. This is an ordered model-fitting of effects and the order matters and should match the underlying reality if the model is to be meaningful. Should the effects in the model be perfectly nested one within the other in a perfect heirarchical and balanced fashion, then this ANOVA structure is both sound and useful. For an unbalanced design (like the one we have), the analysis does go through and sums of squares can be calculated, but there are choices involved that may or may not fit the reality of the data. In our case, clearly OUTLETs and ITEMs fit (nest) within each given PSU. However, neither OUTLETs nor ITEMs nest within one or the other. All ITEMs are not in each OUTLET and just as surely, all OUTLETs are not under each ITEM (or Item-Stratum). The sample design reality is that ITEMs and OUTLETs are independently sampled and then subsequently matched to each other. The best (albeit not perfect) order-fitting of these two effects is ITEM first, then OUTLET. While it seems more natural to think of ITEMs being sampled from some OUTLET, the statistical design reality is that we have a given set of Item-Strata (the actual ITEMs in the model) to which OUTLETs are attached. Our model universe is an AREA by a MAJ_GRP. Each Major Group contains a given set of Item-Strata which can be modeled as the random effect ITEM. The OUTLETs then fit more "naturally" inside these ITEMs.[6] Type I sums of squares are then computed under this ordering. The mean squares for each effect and for error can thus be computed (MS = SS/df). Whenever a given effect is treated in the model as random, then these mean squares are set equal to the Expected Mean Squares, and the solutions to these equations are the estimated variance components. This is one way to estimate the variance components, a strictly model-based approach.

So, what's wrong with using Type I variance components from an ANOVA model? Two things: (1) the ordering of the effects is not a clear-cut decision, indeed cannot be a clear-cut decision unless the data are balanced and nested, and (2) the Type I method can easily enough produce negative components of variance in the effects. If, for instance, MSE > MS(O) a negative variance component for OUTLETs is guaranteed. These negative variance components can then be set to zero (since a variance measure, by definition, cannot be negative), but this is hardly an ideal solution.

5. Variance Components Using REML Estimation

A distributional approach to variance component estimation involves Maximum Likelihood Estimation (MLE), or its counterpart, Restricted Maximum Likelihood (REML) Estimation. We will end up choosing to use REMLs, in fact weighted REMLs, in our final analysis, but a quick look at MLEs in tandem with REMLs can be instructive. Under easily-had

[6] We ran a side analysis on this issue. We computed numerous sets of variance components, first using the PSU-ITEM-OUTLET order for the effects-fitting, then, using the same data, we ran a PSU-OUTLET-ITEM ordering. As a control test, we compared these variance components results with their REML counterparts. The P-I-O ordering matched better to the REML results more than 90% of the time.

regularity conditions, by the multivariate Central Limit theorem, the vector **y** will tend to be distributed multivariate normal.[7] Our model remains:

$$\mathbf{y} = \mathbf{J}\mu + \mathbf{Z}\gamma + \mathbf{e}, \text{ or more simply } \mathbf{y} = \mathbf{J}\mu + \eta,$$

with $\mathbf{y} \sim N(\mu, V(\theta))$, which defines the MLE's likelihood joint density, or
$\eta \sim N(0, V(\theta))$, which defines the REML's likelihood,
with $V(\theta)$ defined below.

Both of these multivariate Normal assumptions are strengthened by a large enough sample size (which we have, perforce, because, in our actual work, $n > 20$ is guaranteed). But the MLE procedure tends to ignore the fact that parameters are being fit (in our case, the grand mean, μ). Since REMLs involve finding maximum likelihood estimates of variance components from the distribution of the residuals η, instead of from the distribution of **y**, this extra parameter μ no longer needs to be fit. Moreover, its degree of freedom gets taken into account under the REML structure, but not under MLE. This is a bias correction,[8] even though neither MLEs nor REMLs are *in se* unbiased estimators. So REMLs give a truer maximum likelihood estimation. Both MLEs and REMLs are order invariant and, as distributional constructs, cannot and so do not allow negative components into their feasibity regions. For these last two reasons, we have chosen maximum likelihood estimation over ANOVA-based estimation for determining our components of variance, and of the two maximum likelihood methods, we have chosen REMLs over MLEs.[9]

The REML Equations

$$\sum_{j=0}^{r} \sigma_i^2 \, tr[Z_j Z_j^T V^{-1} (I - A_0) Z_i Z_i^T V^{-1} (I - A_0)] = Y^T (I - A_0)^T V^{-1} Z_i Z_i^T (I - A_0) Y,$$[10]

$i = 0, \ldots, r$.[11] That's it. We know Y and we know each Z_i. $\mathbf{Z} = [Z_1 \ldots, Z_r]$ is simply the design matrix, $Z_0 = W^{-1/2}$, [12] and A_0 a projection operator that is defined below. V is unknown to begin with, but we make an initial guess with the σ_i^2's, then proceed iteratively to update the V's, which in turn calls for a recalculation of the projection matrix, A_0, and when the σ_i^2's converge, we're done. Simple iteration suffices 90% of the time. And only when one of the variance components is at or very nearly zero does the initial guess much matter. [13] But then, of course, we need to understand where these relatively simple equations come from.

Christensen, in his chapter on Mixed Models and Variance Components,[14] derives these REML equations. The model sets up as above: $\mathbf{y} = \mathbf{J}\mu + \eta$, $\eta \sim N(0, V(\theta))$, $\theta = (\sigma_0^2, \ldots, \sigma_r^2)$, $V(\theta) = \sum_{i=0}^{r} \sigma_i^2 Z_i Z_i^T$, with $Z_0 Z_0^T = W^{-1}$ and $\sigma_0^2 = \sigma_e^2$. Note that the weight structure (W^{-1}), which is a diagonal matrix that replaces the unweighted (I) structure, attaches very simply and naturally to the residual component itself, or alone as an $n \times n$ $Z_0 Z_0^T$ matrix whenever and wherever the algebra calls for it. The REML equations themselves are the partial derivatives of the log likelihood, taken with respect to θ (μ is fitted first,

---

[7] M. Kendall & A. Stuart (1979), *The Advanced Theory of Statistics*, Vol.2 (4th Ed), p. 59

[8] Poduri S.R.S. Rao (1997), *Variance Components Estimation*, p. 99. Rao estimates this negative bias to be $-r \cdot \sigma_e^2 / n$, with r equal to the rank of the design matrix. Should more parameters need to be fit in the model as fixed effects this bias only grows, since r will only increase.

[9] Bayesian Estimation, or Heirarchical Bayes Estimation, could have been tried. For a variety of reasons, we chose not to utilize a Bayesian approach.

[10] These equations can be directly programmed in S-Plus, using simple iteration, and effectively are the equations used by SAS' PROC NLP as well. We used SAS' PROC NLP to calculate all our variance components.

[11] For model (1) r = 3; for model (2) r = 2. This 'r' is the number of random effects in the model. The residual term is σ_o^2.

[12] W is the block diagonal matrix of the $W_{i,j,k}$'s from Section 2.

[13] When all variance components are clearly positive, the choice of initial values is of little import. A vector of 1's will suffice and the S-Plus program will easily converge in less than 10 iterations. When one of the variance components "wants" to be less than zero, SAS employs special algorithms that pull it back into the feasibility region. (We were not able to duplicate SAS's results in this second case in the S-Plus program.) For the third case, when one of the variance components is very close to zero, but still positive, the choice of initial values can be crucial. SAS uses MINQUE initial values, but ANOVA-based variance components can be used just as effectively to achieve the desired quick convergence.

[14] R. Christensen (1987), *Plane Answers to Complex Questions: The Theory of Linear Models*, pp. 235-237.

as a fixed effect, using the usual least squares estimate). The least squares estimate for $\boldsymbol{\eta}$ is then (I – A)Y, where A is some projection operator onto C(**J**). Thus the residuals can be defined up front as (I – A)Y, with the distribution of the residuals written

$$(I - A)Y \sim N(0, (I - A)\, V(\theta)\, (I - A)^T).$$

But the likelihood equation cannot quite yet be derived directly, because this covariance matrix will invariably be singular. The likelihood, however, is a density and a density requires a non-singular covariance matrix in order to exist. (Think of residuals as a null space, living in a world of Lebesgue measure zero.) But if we can linearly transform this $\mathbf{R}^n$ space to an $\mathbf{R}^r$ space then we are in business. Well, we can. We construct an *n x r* matrix (B) of rank *r* such that $C(B) = C(J)^{\perp}$ and $B^T J = 0$.[15] With this B we then transform our distribution to

$$B^T(I - A)Y \sim N(0,\ B^T(I - A)\, V(\theta)\, (I - A)^T B).$$

Then the θ that maximizes $B^T(I - A)Y$ is exactly that same θ that maximizes (I – A)Y. But now we have a measurable density, with its likelihood written thus:

$$f(B^T(I-A)Y \mid \theta) = (2\pi)^{-n/2} \mid B^T(I-A)\, V(\theta)\, (I-A)^T B \mid^{-1/2} \exp[-Y^T(I-A)^T B\, [B^T(I-A)\, V(\theta)\, (I-A)^T B]^{-1} \cdot B^T(I-A)Y / 2].$$

This joint density does exist, since the covariance matrix is now non-singular, and we use this equation to derive our partial derivatives:

$$\partial L / \partial \sigma_i^2 = -\tfrac{1}{2} \operatorname{tr}[(B^T V B)^{-1} B^T Z_i Z_i^T B] + \tfrac{1}{2} Y^T B (B^T V B)^{-1} B^T Z_i Z_i^T B (B^T V B)^{-1} B^T Y.$$[16]

But the likelihood is not dependent on the choice of B.[17] What's more, we can rewrite these partials without using any B term. We know that we can construct $A = A_0 = J (J^T V^{-1} J)^{-1} J^T V^{-1}$. Then $(I - A_0)$ is the projection operator onto C(VB) along C(J).[18] What's more, $(I - A_0) = VB(B^T V B)^{-1} B^T$.[19] This equivalency allows us to rewrite the partial derivatives entirely without a B term.

$$\operatorname{tr}[V^{-1}(I - A_0) Z_i Z_i^T] = Y^T (I - A_0)^T V^{-1} Z_i Z_i^T V^{-1} (I - A_0) Y.$$

Finally, since $(I - A_0)$ is idempotent we can expand the LHS in such a way as to separate out our variance components as proper random variables, while at the same time keeping their respective $\hat{\sigma}_i^2$'s in V. Thus we get

$$\begin{aligned} &\operatorname{tr}[V^{-1}(I - A_0) Z_i Z_i^T] \\ &= \operatorname{tr}[V^{-1}(I - A_0) V V^{-1} (I - A_0) Z_i Z_i^T] \\ &= \operatorname{tr}[V V^{-1} (I - A_0) Z_i Z_i^T V^{-1} (I - A_0)] \\ &= \operatorname{tr}[(\textstyle\sum_{j=0}^{r} \sigma_i^2 Z_j Z_j^T) V^{-1} (I - A_0) Z_i Z_i^T V^{-1} (I - A_0)] \\ &= \textstyle\sum_{j=0}^{r} \sigma_i^2 \operatorname{tr}[Z_j Z_j^T V^{-1} (I - A_0) Z_i Z_i^T V^{-1} (I - A_0)], \end{aligned}$$[20]

and we have derived the working equations for our REML estimation,

$$\textstyle\sum_{j=0}^{r} \sigma_i^2 \operatorname{tr}[Z_j Z_j^T V^{-1} (I - A_0) Z_i Z_i^T V^{-1} (I - A_0)] = Y^T (I - A_0)^T V^{-1} Z_i Z_i^T (I - A_0) Y,$$

and we are done.

Results and Conclusions

Price data from the C&S Archive Database, from mid-1993 through mid-1997, were carefully compiled and massaged, month by month, into one very large dataset. Pricing periods were carefully noted and charted, so that only like price quotes were tracked to each other over time. We then drew on this new database to produce 2-, 6- and 12-month variance components. The 6-month results, for a variety of reasons, were the chosen variance components of record. The 2-month price changes were considered too variable and maintaining a high enough level of sample size over these 12-month periods was considered too difficult. Moreover, no month per cell (AREA by MAJ_GRP) that held fewer than 20 quotes was used in the final results. For the 6-month variance components the average sample size was around 80 quotes for all AREAs except the X's, which averaged around 400 quotes per cell. For the final

[15] To construct B simply calculate a perpendicular projection operator (PPO) onto the null space of J (use (I–A)), and matrix multiply this PPO times an *n x r* matrix of white noise, T. Thus $B^T = T(I-A)$.

[16] See Christensen, p. 231, for the four matrix differentiation results needed to derive these partial derivatives. (See Schott, *Matrix Analysis for Statistics*, pp. 332-333 for complete proof of the fourth matrix differentiation result.)

[17] Ibid, Christensen, p. 234.

[18] Ibid, p. 236.

[19] Ibid, p. 237.

[20] Ibid, p. 237.

variance components we simply averaged all the good results across time (mean # of time periods = 20) per cell. By using PROC MIXED in SAS to produce our REML estimates, we also got standard errors for each of our variance components. The CV's (using mean standard errors over mean variance components) for most of our results were below 1, though these CV's fluctuated somewhat wildly. Some summary statistics on the full table of variance components (over 1500 in all) are presented below:

Table 1. Average Variance Components All MAJ_GRPs – All AREAs

	ITEM	OUTL	PSU Non–A	Resid
AveVC	0.0057	0.0089	0.0113	0.001
AveSE	0.0055	0.0048	0.0022	0.002

Table 2. Average Variance Components by MAJOR_GROUP

MAJOR GROUP	Mean Total Variance
Food – Staples	0.0304
Food – Meats	0.0479
Food – Fruits & Vegetables	0.0689
Food – Other	0.0276
Food Away from Home	0.0158
Entertainment	0.0175
Apparel	0.0399
Household Furnishings	0.0191
Utilities & Fuel	0.0134
Transportation	0.0197
Gasoline	0.0129
Medical	0.0146
Education & Communications	0.0093

7. Acknowledgements

The authors would like to thank Janet Williams for her careful reading of this paper and her helpful comments. In addition, we would like to thank our BLS colleagues, Bob Baskin, Sylvia Leaver and David Swanson for their own earlier work in this area and for their help in reviewing this paper.

8. References

Baskin, R.M. (1995), "Estimation of Variance Components for the U.S. Consumer Price Index", *Proceedings of the Survey Research Methods Section*, American Statistical Association (Vol. 1), 126-131.

Bureau of Labor Statistics, *CPI Detailed Report (Feb 1999)*, Washington, DC, U.S. Government Printing Office.

Christensen, Ronald (1987), *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer-Verlag, New York, NY.

Pfeffermann, D., Skinner, C.J., Goldstein, H., Holmes, D.J., Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multilevel Models (with discussion)", Journal of the Royal Statistical Society, Series B, Volume 60, pp. 23-40.

Kendall, M. & Stuart, A. (1979), *The Advanced Theory of Statistics, Vol 2 (4th ed)*, Macmillan Publishing Co., New York, NY.

Rao, Poduri S.R.S. (1997), *Variance Components: Mixed models, methodologies and applications*, Chapman & Hall, London.

Schott, J.R. (1997), *Matrix Analysis for Statistics*, John Wiley & Sons, New York, NY

Searle, S.R., Casella G., & McCulloch C.E. (1992), *Variance Components*, John Wiley & Sons, New York, NY.