Contents

1 Correlated Errors in Regression

1.1 Introduction

To this point, we have focused on situations where the errors (or equivalently the Y values conditional on the x values) are uncorrelated. We now turn our attention to models allowing for correlation among errors/responses, which can arise in a variety of ways. As before $E(Y_i|\underline{x}_i) = m(\underline{x}_i, \underline{\beta})$ or

$$Y_i|\underline{x}_i = m(\underline{x}_i, \underline{\beta}) + \epsilon_i.$$

Conditional on the x values in the model $\sigma_{ij} = cov(Y_i, Y_j) = cov(\epsilon_i, \epsilon_j)$ is the covariance between Y_i and Y_j and the correlation is $corr(Y_i, Y_j) = \sigma_{ij}/\sigma_i\sigma_j$.

$$Cov(\underline{Y}|\mathbf{x}'s) = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & . & \sigma_{2n} \\ . & . & . & . \\ \sigma_{n1} & \sigma_{n2} & . & \sigma_n^2 \end{bmatrix}.$$

Or $\underline{Y} = \underline{m}(\mathbf{x}, \underline{\beta}) + \underline{\epsilon}$, where

$$\underline{m}(\mathbf{x}, \underline{\beta}) = \begin{bmatrix} m(\mathbf{x}_1; \underline{\beta}) \\ m(\mathbf{x}_2; \underline{\beta}) \\ . \\ m(\mathbf{x}_n; \underline{\beta}) \end{bmatrix}$$

and $Cov(\underline{\epsilon}) = \Sigma$. As with variance terms in our earlier discussion, the σ_{ij} will often be expressed as functions of either the mean or some other parameters. In general this could be written as $\sigma_{ij} = \sigma^2 v_{ij}(\underline{\beta}, \underline{\lambda})$ where the ij allows dependence on other things such as $\underline{x}_i$ and $\underline{x}_j$.

Using least squares

For linear models if we continue to use least squares $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'Y}$, this gives unbiased estimators but $cov(\hat{\boldsymbol{\beta}}) = \Sigma_{\hat{\beta}} = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$ so the naive estimate $MSE(X'X)^{-1}$ is clearly wrong. One way to modify this is to use $\hat{\Sigma}_{\hat{\beta}} = (X'X)^{-1}X'\hat{\Sigma}X(X'X)^{-1}$, where $\hat{\Sigma}$ is some estimate of Σ obtained in a manner particular to the model under consideration. Similar arguments can be made for nonlinear least squares.

Generalized Least squares

Generalized least squares (not to be confused with generalized linear models) estimates the coefficients in a way which takes account of the covariance structure. *If the covariance* Σ *were known*, then it turns out that best linear unbiased estimators (of the coefficients and linear combinations of them) are found via the generalized least squares estimator. For the linear model this is

$$\underline{\hat{\beta}} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{Y}.$$

with $cov(\underline{\hat{\beta}}) = (X'\Sigma^{-1}X)^{-1}$.

The argument for this is that Σ (assumed to be a positive definite matrix; i.e., a nonsingular covariance matrix) can be partition factored as TT' where T is nonsingular. So,

$$\underline{\hat{Y}}^* = T^{-1}\underline{Y} = T^{-1}\underline{m}(\underline{\beta}) + \underline{\epsilon}^*,$$

where $\underline{\epsilon}^* = T^{-1}\underline{\epsilon}$ has covariance I. This reduces the problem to one with constant variance and uncorrelated errors. One can now do least squares on the transformed model. This turns out to be equivalent to **generalized weighted least squares** which minimizes $(\underline{y} - \underline{m}(\underline{\beta}))'\Sigma^{-1}(\underline{y} - \underline{m}(\underline{\beta}))$ as a function of $\underline{\beta}$.

The same argument applies if we only know Σ up to a proportionality constant $\Sigma = \sigma^2 V$, in which case

$$\underline{\hat{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{Y}.$$

with $cov(\underline{\hat{\beta}}) = \sigma^2(X'V^{-1}X)^{-1}$, which would be estimated by $\hat{\sigma}^2(X'V^{-1}X)^{-1}$, where $\hat{\sigma}^2 = MSE$ is the estimated variance from the transformed fit.

Often Σ contains unknown parameters. in which case we would do generalized least squares via

$$\underline{\hat{\beta}} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\underline{Y}$$

Note that the case of $\sigma^2 V$ is a special case of this. This could be two-stage (get $\hat{\Sigma}$ once) or iterative (move back and forth between a new $\hat{\Sigma}$ and new $\hat{\boldsymbol{\beta}}$). An estimate of the covariance of $\hat{\beta}$ is $\hat{\Sigma}_{\hat{\beta}} = (X'\hat{\Sigma}^{-1}X)^{-1}$. This is generally okay asymptotically but for small to moderate samples can be poor since it does not account for uncertainty in $\hat{\Sigma}$.

Under likelihood models we can also consider **maximum likelihood estimation** of the mean and covariance parameters simultaneously. For some models this is equivalent to iteratively "reweighted" generalized least squares.

1.2 A single regression over time

Suppose t indexes time order and we have a model $Y_t|\underline{x}_t = m(\underline{x}_t, \underline{\beta}) + \epsilon_t$. (We will not consider dynamic models where x can contain previous values of y.) Many designed experiments with the regressors (x's) controlled by the experimenter must be run over time, often with only one observation at a time. Here the definition of the expected value and variance of $Y \mid x$ is defined in terms of the expected behavior of the response in the presence of x over some specified time frame of interest (the time over which the experiment is run). There may be some random quantities (such as unsuspected environmental factors) over time that influence "errors" over time. This creates correlated errors over time.

We will illustrate the problem and remedies in the context of **a first order autoregressive, AR(1), model for the errors** with observations equally spaced over time, but the general approach extends to more complicated models on the errors. The model is

$$\epsilon_t = \rho\epsilon_{t-1} + u_t,$$

where the u_t are assumed to be independent and identically distributed with mean 0 and variance σ_u^2. This is an autoregressive 1 model for serial correlation. If you consider stretching back infinitely in time, then it can be shown that $\epsilon_t = \sum_{s=0}^{\infty} \rho^s u_{t-s}$ so $E(\epsilon_t) = 0$,

$$V(\epsilon_t) = \sigma_\epsilon^2 = \sum_{s=0}^{\infty} \rho^{2s} V(u_{t-s}) = \sigma_u^2/(1-\rho^2), cov(\epsilon_t, \epsilon_{t+k}) = \rho^{|k|}\sigma_\epsilon^2 \text{ and } corr(\epsilon_t, \epsilon_{t+k}) = \rho^{|k|}.$$

$$\Sigma = Cov(\underline{Y}) = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & . & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & . & \rho^{n-2} \\ . & . & . & . & . & . \\ \rho^{n-1} & \rho^{n-2} & . & . & . & 1 \end{bmatrix}.$$

Consider a least squares fit (which still provides unbiased estimators if the model is correct) with residuals $r_1, \ldots, r_n$. A plot of the residuals versus time is helpful in detecting serial correlation over time .

The Durbin-Watson test, based on the test statistic $D = \sum_{t=1}^{n}(r_t - r_{t-1})^2 / \sum_{t=1}^{n} r_t^2$, provides a test of $H_0 : \rho = 0$ under assumption that the u_t are normally distributed. Details of critical values for the test can be found in most linear regression books (and P-values can be obtained in many software packages). The test is based on normality and an alternative is to carry out a bootstrap test. Some procedures will test for higher order lag correlations as will be discussed in an example.

A commonly used estimate of ρ is

$$\hat{\rho} = \sum_{t=2}^{n} r_t r_{t-1} / \sum_{t=2}^{n} r_{t-1}^2,$$

which is based on a linear regression with no intercept treating the r_t as if they are the ϵ_t. There are alternate estimators but this is the most commonly used one. This $\hat{\rho}$ can now be used to get generalized least squares estimates.

- Yule- Walker/ GLS approach.

With $\Sigma = \sigma^2 V$ as above then

$$\hat{V} = \begin{bmatrix} 1 & \hat{\rho} & \hat{\rho}^2 & \hat{\rho}^3 & . & \hat{\rho}^{n-1} \\ \hat{\rho} & 1 & \hat{\rho} & \hat{\rho}^2 & . & \hat{\rho}^{n-2} \\ . & . & . & . & . & . \\ \hat{\rho}^{n-1} & \hat{\rho}^{n-2} & . & . & . & 1 \end{bmatrix}$$

and the GLS estimator (called the Yule-Walker approach in this context) is

$$\underline{\hat{\beta}}_{yw} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}\underline{Y}.$$

with $\hat{\Sigma}_{\hat{\beta}} = MSE(X'\hat{V}^{-1}X)^{-1}$, where MSE is the estimate of σ^2 from the transformed model.

For the AR(1) structure it can be shown explicitly that $\hat{V} = TT'$ with

$$T^{-1} = \begin{bmatrix} (1-\hat{\rho}^2)^{1/2} & 0 & 0 & 0 & ... & 0 \\ -\hat{\rho} & 1 & 0 & 0 & ... & 0 \\ 0 & -\hat{\rho} & 1 & 0 & ... & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & ... & -\hat{\rho} & 1 \end{bmatrix}.$$

This means one can calculate explicitly $\hat{\underline{Y}}^* = T^{-1}\underline{Y}$ and $T^{-1}\underline{m}(\underline{x}, \underline{\beta})$ for use in a transformed model $\hat{\underline{Y}}^* = T^{-1}\underline{m}(\underline{x}, \underline{\beta}) + \underline{\epsilon}^*$ with $cov(\underline{\epsilon}^*)$ taken as $\sigma^2 I$.

For linear regression, the transformed model is $Y_t^* = \underline{x}_t^{*'}\underline{\beta} + u_t^*$, where

$Y_1^* = (1-\hat{\rho}^2)^{1/2} Y_1$ and $\underline{x}_1^* = (1-\hat{\rho}^2)^{1/2}\underline{x}_1$

$Y_t^* = Y_t - \hat{\rho} Y_{t-1}$ and $\underline{x}_t^* = \underline{x}_t - \hat{\rho}(\underline{x}_{t-1})$, $t = 2, \ldots n$,

Also, notice that directly that for the linear model, for $t = 2, \ldots, n$, $Y_t^* = Y_t - \hat{\rho} Y_{t-1} = (\underline{x}_t'\underline{\beta} + \epsilon_t) - \hat{\rho}(\underline{x}_{t-1}'\underline{\beta} + \epsilon_{t-1}) = (\underline{x}_t' - \hat{\rho}\underline{x}_{t-1}')\underline{\beta} + \epsilon_t - \hat{\rho}\epsilon_{t-1} = \underline{x}_t^{*'}\underline{\beta} + \hat{u}_t$ where $\underline{x}_t^* = \underline{x}_t - \hat{\rho}\underline{x}_{t-1}$ and $\hat{u}_t = \epsilon_t - \hat{\rho}\epsilon_{t-1}$, which if $\hat{\rho}$ were ρ would be exactly u_t. This suggests the same type of transformation resulting from the YW/GLS method for observations 2 to n. Some will fit the transformed model using only these $n-1$ values; this sometimes referred to as the Cochrane-Orcutt procedure..

- If autocorrelation is still present, the process can be repeated.

- As with our earlier weighted least squares, the standard errors that are given are generally underestimates since they do not account for uncertainty arising from estimation of ρ. This can be important and the standard errors/covariance matrix for the coefficients can be evaulated via the bootstrap.

- Maximum Likelihood under normality.

Here the u_t are assumed i.i.d. $N(0, \sigma^2)$ and the likelihood is maximized in all parameters. The standard errors for the coefficients do account for uncertainty from estimating ρ since we are using a full maximum likelihood approach, but they depend on the normality assumption. But, all inferences are based on asymptotic results that may be questionable with small to moderate sample sizes.

Additional Remarks:

- The error model can be extended to involve more lags, that is autoregressive models of higher order. This is easily handled in proc autoreg.

- The model can be extended to allow for heteroscedasticity also. Proc autoreg will test automatically for a particular type of heteroscedasticity arising from what is known as a GARCH model.

- Another type of heteroscedasticity is where u_t has a variance which is a function of the x's and/or time. This can be checked in much the way that we did for uncorrelated situations by using a fitted value for u_t, say $\hat{u}_t$ and plotting it in various ways. The residual option in autoreg gives estimates of the u_t's.

- There are a number of other models, tests, etc. that can be entertained. See the documentation on proc autoreg (which is part of the SAS ETS module) and either Judge et. al., Griffiths et al. , or other Econometrics books in particular.

- As noted for a non-linear model, one can still transform in a similar way but the transformation is not as clean without the linearity.

References

- Kutner et al. Chapter 12. We've covered hear much of what is in their 12.1-12.3 and the Cochrane-Orcutt procedure in 12.4.

- Griffiths et al. Chapter 16.

- Many other Econometrics books address the topic of serial correlation.

1.3 Correlated errors in repeated measures/clustered data: Linear Models

In many regression problems, the n observations can be divided up according to whether they were made on some "main unit". The different observations on a main unit are often called repeated measures. Our focus

here is when these repeated measures involve the response variable (and perhaps some of the predictors) measured on different "occasions" (which might mean different times, different spatial locations, etc.).

EXAMPLES

1. For each of 20 boys have measure of length of ramus bone at 8, 8.5, 9 and 9.5 years of age. Interest is in model for change over time that applies to the population of boys and perhaps understand how models change across boys.

2. Have a number of chicks assigned to four treatment groups. Observe weight of chicks over 21 days. Objective is a model for weight gain over time and comparison of the regression curves across the four treatments.

3. Have four fields. Each field divided up into 6 subplots. Each subplot receives some combination of x_1 = delay in cultivation (3 or 10) and x_2 = lbs of hydroxide per acre. Response is number of quackgrass shoots per square acre. The field is the main unit.

4. Investment. Have three firms (GM, GE and Westinghouse). Have yearly data on x_1 = gross investment, x_2 = market value of firm at end of the previous year and y = value of stock of plant and equipment at end of previous year.

Assume there are N "units" in the sample and that unit u in the sample has T_u measurements, denoted

$$\underline{Y}_u = \begin{bmatrix} Y_{u1} \\ . \\ . \\ Y_{uT_u} \end{bmatrix}.$$

The u here is referring to the position in the sample not to a specific individual unit. The actual population unit ending up in this position may be random. This will become clear later when we describe random effects.

BALANCED CASE refers to all $T_u = T$.

Assume

$$E(Y_{uj}) = \mu_{uj} = x_{uj1}\beta_1 + x_{uj2}\beta_2 + \ldots + x_{ujp}\beta_p = \underline{x}'_{uj}\underline{\beta},$$

where the x values are fixed (either the x's are actually fixed or are realizations of random quantities and are viewing the model as a conditional one) and $\beta_1, \ldots, \beta_p$ are unknown parameters. The β's are the "regression parameters" which distinguishes them from parameters that enter into the covariance structure.

$$E(\underline{Y}_u|X_u) = \begin{bmatrix} x_{u11} & . & x_{u1p} \\ . & . & . \\ x_{uT1} & . & x_{uTp} \end{bmatrix} \underline{\beta} = X_u\underline{\beta},$$

with $\underline{\beta}' = (\beta_1, \ldots, \beta_p)$.

The X_u can contain information about the main unit (such as sex, age, a treatment applied to the unit for the whole experiment, etc.) and/or within unit factors/variables (e.g., treatments applied at repeated measures occasions, the time points in longitudinal settings, time varying covariates, etc.). The X_u might also involve variables created using dummy variables to allow different coefficients per unit if we are treating the units as fixed.

The covariance structure is specified by

$$cov(\underline{Y}_u) = \Sigma_u \qquad\qquad cov(\underline{Y}_u, \underline{Y}_w) = \Sigma_{uw}, u \neq w.$$

When there is no correlation among units, $\Sigma_{uw} = 0$ for $u \neq w$. Combining all of the observations together

$$\underline{Y} = \begin{bmatrix} \underline{Y}_1 \\ . \\ . \\ \underline{Y}_N \end{bmatrix},$$

$$E(\underline{Y}) = \begin{bmatrix} X_1 \\ . \\ . \\ X_N \end{bmatrix} \underline{\beta} = X\underline{\beta}, \;\; Cov(\underline{Y}) = \Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_{12} & \Sigma_{13} & . & \Sigma_{1N} \\ \Sigma_{21} & \Sigma_2 & \Sigma_{23} & . & \Sigma_{2N} \\ . & . & . & . & . \\ \Sigma_{N1} & . & . & \Sigma_{N-1,N} & \Sigma_N \end{bmatrix}.$$

1.3.1 Modeling the Covariance structure

The correlation can arise for various reasons including:

- correlation among observations over time (or space) within a unit even if we consider the unit as fixed.

- correlation arising from random unit effects.

- correlation among observations over different units, often arising from random effects at a point in time that influence all of the units.

This leads to how to model the covariance matrices? There is usually some simplifying structure on Σ as it cannot be left completely unspecified. There needs to be some shared parameters in the different covariance matrices entering Σ. There are a variety of ways to model the covariance structure. Here we focus on models that are most appropriate when there is no correlation among subjects and later consider models with correlation among different "subjects".

Modelling the variance/covariance structure of the random terms is important from a couple of perspectives. Firstly, if the covariance structure is incorrectly specified then the analysis on the mean parameters is often incorrect. If too general a model for covariance is used, the estimate for $\underline{\beta}$ may be inefficient since there are more terms in the covariance matrix than needed. Sometimes the covariance structure and estimation of the parameters that enter into it may be of interest in its own right.

A random coefficients example

Suppose the repeated measures are over time and the *uth* unit in the sample will be observed at $x_{u1}, \ldots, x_{uT_u}$ (these can be time points, doses, etc.). Consider a population of units and let β_{k0} and β_{k1} denote the coefficients (intercept and slope respectively) for a simple linear regression model over time, for the kth unit *in the population*. That is, we are assuming that for this particular unit over time the expected value at x is $\beta_{k0} + \beta_{k1}x$. Notice that we are allowing each unit to have their own coefficients. Assume the population mean coefficients are β_0 and β_1 (these are average coefficients across the population of units) with a covariance matrix

$$G = \begin{bmatrix} v_{b0} & v_{b0,b1} \\ v_{b0,b1} & v_{b1} \end{bmatrix}.$$

Now, assume that a random sample of n units is chosen and let (b_{u0}, b_{u1}) denote the random coefficients for the uth selected unit in the sample. If the uth selected unit in the sample is the kth unit in the population, then $b_{u0} = \beta_{k0}$ and $b_{u1} = \beta_{k1}$. Under the assumption of random sampling (assuming the population of units is large)

$$E\begin{bmatrix} b_{u0} \\ b_{u1} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \text{and } Cov\begin{bmatrix} b_{u0} \\ b_{u1} \end{bmatrix} = G.$$

Conditionally, given the selected units and hence given b_{u0} and b_{u1}, $Y_{uj} = b_{u0} + b_{u1}t_{uj} + \epsilon_{uj}$, and unconditionally (over random selection of units)

$$Y_{uj} = \beta_0 + \beta_1 x_{uj} + (b_{u0} - \beta_0) + (b_{u1} - \beta_1)x_{uj} + \epsilon_{uj}.$$

This can be expressed as

$$Y_{uj} = \underline{x}'_{uj}\underline{\beta} + \underline{z}'_{uj}\underline{v}_u + \epsilon_{uj},$$

where $\underline{x}'_{uj} = (1, x_{uj})$, $\underline{z}'_{uj} = (1, x_{uj})$ (the same as $\underline{x}'_{uj}$ in this case), $\underline{\beta}' = (\beta_0, \beta_1)$ and

$$\underline{v}_u = \begin{bmatrix} b_{u0} - \beta_0 \\ b_{u1} - \beta_1 \end{bmatrix}$$

contains random effects with expected value $\underline{0}$ and covariance matrix G.

Hence, in this case the random effects in v_u enter due to the random selection of units via the random coefficients. The overall model is

$$\underline{Y}_u = X_u\underline{\beta} + Z_u\underline{v}_u + \underline{\epsilon}_u,$$

where

$$X_u = Z_u = \begin{bmatrix} 1 & x_{1u} \\ 1 & x_{2u} \\ . & . \\ 1 & x_{T_u} \end{bmatrix}.$$

is a fixed matrix which incorporates the regression model being used. In this setting, where each of the coefficients in the fixed effect part of the model is random, X_u and Z_u are the same.

General two-stage model.

Generalizing along the lines of the development of the example above, a general mixed/two-stage model used for modeling the covariance structure is

$$\underline{Y}_u = X_u\underline{\beta} + Z_u\underline{v}_u + \underline{\epsilon}_u,$$

- X_u is known. The model is conditional on X_u's.

- $\underline{\beta}$ is a fixed, but unknown set of parameters

- Z_u is a known matrix with T_u rows and q columns of rank q. Note that we are allowing the number of observations on the *uth* unit to be T_u and hence change with u.

- The vector $\underline{v}_u$ is a vector of random effects associated with the uth unit in the sample and is assumed to have mean $\underline{0}$ and covariance G.

- $\underline{\epsilon}_u$ (which contains remaining noise/error remaining after accounting for the fixed part and the random effects captured in $\underline{v}_u$) has $E(\underline{\epsilon}_u) = \underline{0}$ and $\Sigma_{\epsilon u} = Cov(\underline{\epsilon}_u)$.

This leads to:

$$Cov(\underline{Y}_u) = \Sigma_u = Z_u G Z'_u + \Sigma_{\epsilon u}. \tag{1}$$

If $\Sigma_{\epsilon u}$ is allowed to be unstructured then nothing is gained by this model since Σ_u is also unstructured. So, the two stage approach is only useful when some simplifying assumptions about $\Sigma_{\epsilon u}$ are made. One assumption often made is referred to as **conditional independence** in which $\Sigma_{\epsilon u} = \sigma^2_u I$, where I is the identity matrix of size T_u. This says that conditional on the random unit effects we have a standard linear model with the errors being uncorrelated with a common variance. The result is

$$Cov(\underline{Y}_u) = \Sigma_u = Z_u G Z'_u + \sigma^2_u I. \tag{2}$$

In this case Σ_u has (q(q+1)/2) + 1 distinct parameters; the q(q+1)/2 arising from the number of parameters in G and the 1 accounting for the σ_u^2 in $\Sigma_{\epsilon u}$.

An alternative which is considered when the repeated measures are over time and time is equally spaced is to use an AR(1) covariance matrix for $\Sigma_{\epsilon u}$.

Types of covariance matrices

Here we describe some general structures for the various covariance matrices involved.

- UNCORRELATED/CONSTANT VARIANCE STRUCTURE

  $\sigma^2 I$. (Constant variance, no correlation). Often used for Σ_ϵ. If this were used for the overall Σ then we are back to standard linear regression.

- GENERAL UNSTRUCTURED COVARIANCE

  An unstructured or general covariance matrix places no restrictions on the covariance matrix except the usual ones of symmetry and positive definiteness which qualify it as a valid covariance matrix. If the matrix is $T \times T$, then there are T(T+1)/2 distinct parameters in it. This might be used for the Σ_u for a subject or for the covariance matrix G associated with the random effects piece $\underline{v}_u$ in the two-stage model.

- COMPOUND SYMMETRY

  The covariance matrix is said to be COMPOUND SYMMETRIC if it is of the form

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

  or equivalently

$$\begin{bmatrix} \sigma_1^2+\sigma_2^2 & \sigma_2^2 & \dots & \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2+\sigma_2^2 & \dots & \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 & \dots & \sigma_2^2 & \sigma_1^2+\sigma_2^2 \end{bmatrix}.$$

  If such a covariance was used to represent Σ_u, then $\sigma^2 = \sigma_1^2+\sigma_2^2 = V(Y_{uj})$ and for $j \neq j', cov(Y_{uj}, Y_{uj'}) = \rho\sigma^2 = \sigma_2^2$ or $corr(Y_{uj}, Y_{uj'}) = \rho$.

  There are two distinct parameters, σ^2 and ρ in the first parameterization, and σ_1^2 and σ_2^2 in the second version. The second version is often used since σ_1^2 and σ_2^2 represent variance components due to two different random effects.

  Compound symmetry arises naturally in certain random effects models. This is a special case of the two-stage models above. Suppose that

$$Z_u\underline{v}_u = \begin{bmatrix} 1 \\ 1 \\ . \\ 1 \end{bmatrix} v_u$$

  so $\underline{v}_u = v_u$ (a univariate random variable) which represents a random intercept term associated with selection of the uth unit and $Z_u = \underline{1}$, a vector of 1's. If $Var(v_u) = \sigma_2^2$ and $\Sigma_{\epsilon_u} = \sigma^2 I$, then $Cov(\underline{Y}_u) = \Sigma_u = \underline{1}\sigma_2^2\underline{1}' + \sigma^2 I$ leading to the compound symmetric structure above with $\sigma^2 = \sigma_1^2$.

- AUTOREGRESSIVE(1) covariance structure is

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho^{T-2}\sigma^2 & \rho^{T-1}\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho^{T-3}\sigma^2 & \rho^{T-2}\sigma^2 \\ . & & \dots & & . \\ \rho^{T-1}\sigma^2 & \rho^{T-2}\sigma^2 & \dots & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

where $\sigma^2 = V(Y_{uj})$ and $cov(Y_{uj}, Y_{uj'}) = \sigma^2\rho^{|j-j'|}$.

This is often used as a model for Σ_ϵ, the within unit covariance matrix for observations observed over **equally spaced time points.** This covariance structure also has two distinct parameters in it.

1.3.2 Inferences with no among "subject" correlations

Here we assume $\Sigma_{uw} = 0$ for $u \neq w$, so the model assumptions are:

A1. $E(\underline{Y}_u) = \underline{\mu}_u = X_u\underline{\beta}$

A2. $cov(\underline{Y}_u) = \Sigma_u$ where Σ_u depends on some underlying set of parameters with the number of parameters involved and the notation for them depending on the particular covariance structure assumed (unstructured, compound symmetry, etc.). Denote this collection of parameters by $\underline{\phi}$. So, for example, when $\Sigma_u = \Sigma^*$ is unstructured $\underline{\phi}$ consists of $\sigma_{11}, \dots, \sigma_{1T}, \sigma_{22}, \sigma_{23}, \dots, \sigma_{2T}, \dots, \sigma_{TT}$, while for compound symmetry in Σ^*, $\underline{\phi}$ consists of σ_1^2 and σ_2^2 (or σ^2 and ρ).

A3. $\underline{Y}_1, \underline{Y}_2, \dots \underline{Y}_N$ are uncorrelated. $Cov(\underline{Y}_u, \underline{Y}_w) = 0$ for $u \neq w$.

For distribution/likelihood based analyses, assume in addition

A4. $\underline{Y}_u$ has a multivariate normal distribution (with mean $X_u\underline{\beta}$ and covariance matrix Σ_u). (In this case the uncorrelated assumption implies independence).

In a **balanced design with constant covariance**, each unit is observed at T repeated levels ($T_u = T$ for each u) and Σ_u is constant over u. For this case, denote the common covariance by Σ^*.

$$\Sigma^* = Cov(\underline{Y}_u) = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1T} \\ . & \dots & . \\ \sigma_{T1} & \dots & \sigma_{TT} \end{bmatrix}.$$

$\sigma_{jj} = V(Y_{uj})$, $cov(Y_{uj}, Y_{uj'}) = \sigma_{jj'} = \sigma_{j'j}$.

- **Generalized Least squares for $\underline{\beta}$** Since Σ is block diagonal, the generalized least squares estimate is

$$\underline{\hat{\beta}} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\underline{Y} = \left[\sum_{u=1}^{n} X_u'\hat{\Sigma}_u^{-1}X_u\right]^{-1}(\sum_{u=1}^{n} X_u'\hat{\Sigma}_u^{-1}\underline{Y}_u).$$

In general the estimator $\underline{\hat{\beta}}$ is asymptotically multivariate normal with mean $\underline{\beta}$ and a covariance matrix

$$\Sigma_{\hat{\beta}} = (X'\Sigma^{-1}X)^{-1} = \left[\sum_{u=1}^{N} X_u'\Sigma_u^{-1}X_u\right]^{-1},$$

which can be estimated by

$$\hat{\Sigma}_{\hat{\beta}} = \left[\sum_{u=1}^{N} X_u'\hat{\Sigma}_u^{-1}X_u\right]^{-1}.$$

This estimate is correct if we have the variance/covariance model right. A robust estimator, which is similar in spirit to White's estimator uses

$$\hat{\Sigma}_{\hat{\beta},robust} = B\left[\sum_{u=1}^{N} X_u'\hat{\Sigma}_u^{-1}\underline{e}_u\underline{e}_u'\hat{\Sigma}_u^{-1}X_u\right]B,$$

where $B = \left[\sum_{u=1}^{N} X_u'\hat{\Sigma}_u^{-1}X_u\right]^{-1}$ and $\underline{e}_u = \underline{Y}_u - X_u\hat{\underline{\beta}}$. (In SAS proc mixed this robust estimator can be obtained using the empirical option in the proc mixed statement.)

SPECIAL CASE: If you just use simple least squares ignoring anything about the variance- covariance structure then the robust estimator of the covariance of $\hat{\underline{\beta}}$ is

$$\hat{\Sigma}_{\hat{\beta},robust} = (X'X)^{-1}\left[\sum_{u=1}^{N} X_u'\underline{e}_u\underline{e}_u'X_u\right](X'X)^{-1},$$

where X is the overall design matrix.

Under normality ,with the right $\hat{\Sigma}_u$'s the gls estimator above is the MLE for $\underline{\beta}$ and the estimate $\hat{\Sigma}_{\hat{\beta}}$ agrees with the estimated covariance from the expected information approach.

We need to be able to estimate the "variance parameters" in $\underline{\phi}$ (in order to estimate the Σ_u). In some problems we can get unbiased estimates of the ϕ's in closed form. Otherwise there are some estimating equations which yield estimates of the ϕ's which arise from likelihood considerations under normality (see below) or some other criterion (one is Minimum Variance Quadratic Unbiased Estimation or MIVQUE0 in SAS). If the variance parameters are estimated just once, then the estimate of $\underline{\beta}$ is two-stage generalized least squares. Most of the mixed model software will iterate though, updating both the estimate of $\underline{\beta}$ and the estimate of $\underline{\phi}$.

It turns out the estimates of ϕ's arising from a likelihood motivation under normality, are consistent estimates under just the moment assumptions, so the inferences on $\underline{\beta}$ are okay without normality.

- **Likelihood approaches.**

  The discussion in this section is in the context of uncorrelated $\underline{Y}_u$ where some specialized expressions occur but many of the ideas are readily extended to allow among correlation structure.

  Distribution based methods rely on specific assumptions about the probability distribution of the random data. For our settings, this will be restricted to making normality assumptions. Distribution free or nonparametric methods here means proceeding just with assumptions about the mean and covariance model as given above.

Maximum Likelihood Estimation

Under multivariate normality, the joint density of all of the data can be written as

$$f(\underline{y}_1, ..., \underline{y}_N \mid \beta, \underline{\phi}) = f_{\underline{Y}_1}(\underline{y}_1 \mid \underline{\beta}, \underline{\phi}) \cdot \ldots \cdot f_{\underline{Y}_N}(\underline{y}_N \mid \underline{\beta}, \underline{\phi}).$$

where $f_{\underline{Y}_u}(\underline{y}_u \mid \underline{\beta}, \underline{\phi})$ is the multivariate normal density function with mean $X_u\underline{\beta}$ and covariance matrix Σ_u (which involves $\underline{\phi}$). The density function has the y's as the arguments for given fixed $\underline{\beta}$ and $\underline{\phi}$.

The <u>likelihood function</u> $L(\underline{\beta}, \underline{\phi} \mid \underline{y}_1, ...\underline{y}_N)$ is simply $f(\underline{y}_1, ..., \underline{y}_N \mid \beta, \underline{\phi})$ but viewed as a function of the parameters with $\underline{y}_1, \ldots \underline{y}_N$ the fixed observed values in the sample.

The <u>Maximum likelihood estimates</u>: (MLE's) of the parameters, denoted $\hat{\underline{\beta}}$ and $\hat{\underline{\phi}}$ are the values of β and $\underline{\phi}$ which maximize the likelihood function over the parameter space, which consists of acceptable

values of $\underline{\beta}$ and $\underline{\phi}$. In many of our problems there is a unique point which maximizes the likelihood. In general, one must watch out for the occurrence of local maxima and more importantly for cases where no maximum occurs over the acceptable range of the parameter space. For the problems of interest to us here, the latter can be of concern. First, some of the components in $\underline{\phi}$ are variances which carry the restriction of being nonnegative. In addition, the resulting estimate of each Σ_u must be positive definite.

Restricted Maximum likelihood Estimation

An alternative is to use restricted maximum likelihood estimators (REML estimates). These estimators are really geared towards improved estimation of the parameters entering into the covariance matrix. The details can get rather technical, so we give just a very general idea and refer the interested reader to Hocking (1985, p. 244) Christiansen (1987) and references therein.

As noted by many, the name is a misnomer in that the procedure does not involve restrictions on the parameters. Rather the REML method uses a modified likelihood which addresses the variance parameters in a different way. One way to view REML estimation is the following. First the ordinarily least squares estimate for $\underline{\beta}$, ($\underline{b}_L$ say) is found and then the likelihood function of the residuals ($\underline{Y}_u - X_u\underline{b}_L$), which depends on $\underline{\phi}$, is maximized to obtain estimates for the components of the covariance matrix. After this is done one can then update estimate of $\underline{\beta}$ using generalized weighted least squares with the estimated covariance matrix and this could be iterated. Essentially the REML estimates maximize a modified likelihood.

An advantage of REML is that in case of balanced data it usually produces estimators of the elements of the covariance matrix which are unbiased (and in other senses optimal). For example if we apply REML in the standard regression model with constant variance and uncorrelated errors, we get $MSE = SSE/(n-p)$, which is an unbiased estimator of σ^2.

General comments on inference

- For either ML or REML, the estimate of $\underline{\beta}$ is of the form

$$\left[\sum_{u=1}^{N} X_u'\hat{\Sigma}_u^{-1}X_u\right]^{-1}\left(\sum_{u=1}^{N} X_u'\hat{\Sigma}_u^{-1}\underline{Y}_u\right)$$

  where $\hat{\Sigma}_u$ (obtained through the estimates of the components in $\underline{\phi}$) is the estimate of Σ_u. This agrees with generalized least squares.

- In the balanced cases where $\Sigma_u = \Sigma^*$ is compound symmetric or unstructured, closed form non-iterative solutions sometimes exist for both ML and REML estimates and exact confidence intervals and tests are available under normality.

- In many cases, obtaining either the ML or REML estimates requires iterative techniques to maximize the necessary functions (equivalently solve the associated estimating equations). There are a number of methods, including the Newton-Raphson method , Fisher's scoring method, the EM algorithm and modifications of these.

- In some cases, the estimate of $\underline{\beta}$ is the same using either the ML or REML approach. However the estimates of $\underline{\phi}$ will differ which will in term change the estimated covariance of the regression parameters (and hence the standard errors which go with the regression coefficients).

- Asymptotically, the ML and REML estimators of $\underline{\phi}$ are equivalent. For large samples, there often is not much difference between the two.

- There are some cases in which under the assumption of normality, exact distributional results are available for estimators of either the regression parameters or the covariance parameters.
- Under distributional assumptions (as used here), there are two common approaches to estimating the covariance parameters of the estimated parameters. One uses what is called the **expected information matrix** the other uses the **observed/empirical information matrix.** (For REML it is based on a modified log-likelihood). For some complete data problems the two are equivalent. In many cases they are not and there is no clear consensus on which is preferred, though there are some arguments for observed information in missing data problems. (Efron and Hinkley).

  In proc mixed, the estimated covariance matrix associated with the estimate of ϕ parameters usually corresponds to an observed information approach (in that it uses the Hessian matrix associated with the objective function, whether it be via ML or REML), but there are some exceptions. Related inferences for the ϕ parameters in proc mixed are dependent on the normality assumption.

  The estimated covariance for the regression parameters agrees with the expected information approach and these are robust to the normality assumption.
- Confidence intervals and Testing.

As noted above, for some of the complete data cases with particular covariance structures and multivariate normality, there are exact confidence intervals and tests available. In general , approximate tests and confidence intervals follow procedures similar to those we have used (normal based intervals, Wald tests, likelihood ratio tests, bootstrapping, etc.).

Recall that the **Wald statistic** for testing $H_0 : H\underline{\beta} = \underline{h}$ (H is $q \times p$ of rank q) is $C_W = (H\underline{\hat{\beta}} - \underline{h})'(H\hat{\Sigma}_{\hat{\beta}}H')^{-1}(H\underline{\hat{\beta}} - \underline{h})$ which is approximately chi-square with q degrees of freedom under H_0.

SAS proc mixed treats W/q as approximately $F(q, df)$ with a particular (sometimes exact, sometimes approximate) way to compute df. This same approach can be taken to test linear hypotheses about the parameters in the covariance matrix Wald-type tests may not work very well for small to moderate sample sizes, especially with lots of parameters present.

Likelihood ratio tests

Consider two models M1 and M2 where M1 is nested in M2. That is, M2 is a general model of which M1 is a special case.

Example 1: Consider a polynomial trend model. Suppose model M2 uses a quadratic model $E(Y_{uj}) = \beta_0 + \beta_1 t_j + \beta_2 t_j^2$ while model M1 has a linear model $E(Y_{uj}) = \beta_0 + \beta_1 t_j$ and the same covariance structure is present in each model. Note that $M1$ is a special case of $M2$ obtained by setting $\beta_2 = 0$. It is crucial that the same covariance structure is assumed for each model.

Example 2: Model M2 has $E(\underline{Y}_u) = \underline{\mu}_u$ and $cov(\underline{Y}_u) = \Sigma^*$ where Σ^* is unstructured. Model M1 has the same mean structure but Σ^* is compound symmetric.

We wish to test H_0 : model is M1 versus H_A : model is M2. In example 1, this amounts to assuming the quadratic model is correct to start and we wish to test the null hypothesis that $\beta_2 = 0$. In example 2, this would mean testing the hypothesis that the Σ^*, assumed unstructured to start, is compound symmetric. Let L_1 = value of maximized log likelihood under model M1 and L_2 = value of maximized log likelihood under model M2. The **likelihood ratio test statistic** is $C_{lr} = 2*(L_2 - L_1) = -2L_1 - (-2L_2)$. For suitable sample sizes, C is approximately chi-square with q degrees of freedom where q = number of distinct parameters in model M2 - number of distinct parameters in model M1. The test rejects H_0 if C is bigger than $\chi^2(\alpha, q)$.

FOR TEST ABOUT THE COVARIANCE STRUCTURE YOU CAN USE THE LOG LIKELIHOOD VALUES FROM EITHER REML OR ML (WITH THE SAME MEAN MODEL IN BOTH FITS).

IF LIKELIHOOD RATIO TESTS ARE TO BE CONSTRUCTED FOR FIXED EFFECTS THEY

MUST BE DONE USING ML AND NOT REML SOLUTIONS! WITH THE SAME COVARIANCE MODEL IN BOTH FITS.

Unbalanced/ missing data settings

Unequal sample sizes arises through unbalanced or missing data. One way to model this is to consider that there are T potential repeated levels at which a unit could be observed. Unit u is observed at $T_u \leq T$ of these levels. Think of Σ^* in the balanced complete case as the covariance matrix for the vector of observations on unit u if observed at all T points. This can be modelled in the various ways already given. Can take Σ_u to be composed of the T_u rows and columns of Σ^* corresponding to the T_u levels at which data was obtained.

NOTE!! Handling missing data in this way is only correct under certain assumptions about the missing data mechanism, namely that it is what is called MCAR or MAR (missing completely at random or missing at random); See Little and Rubin's "The Analysis of Missing Data". In MCAR the probability of an observation missing is constant. In MAR the probability of an observation missing can depend on observed values but not depend on the unobserved missing values. In other settings the missingness is said to be informative and other techniques are needed.

1.4 Linear models having within and among "subject" correlation including time-series/cross-sectional data

It is possible to have the variance/covariance structure coming from two "directions". For example, we might have units designated as in the discussion of the preceding section (indexed by u; in most of the previous section the units were random, but they might also have been fixed and the correlation among observations came from the "within unit" behavior) but there might be random effects corresponding to "columns" which cut across the "units". For example, with the columns indicating time we might have a random time effect added to the model. The random time effect at a time influence all observations occuring at that time. More generally one might want to allow a general covariance structure on the errors at a particular time. As another example, suppose we have observations on different individuals (ids) arising from the use of different machines. There can be random id effects and random machine effects.

If Y_{uj} is the jt observation on the uth unit we can write as before:

$$Y_{uj} = \underline{x}'_{uj}\underline{\beta} + \delta_{uj} \tag{3}$$

where $E(\delta_{uj}) = 0$.

In the preceding section $\delta_{uj} = \underline{z}'_{uj}\underline{v}_u + \epsilon_{uj}$ with observations from different u's assumed uncorrelated and $Cov(\underline{\epsilon}_u) = R_u$ (maybe with some structure, mabye not).

It can get complicated to write down the fullest generalization of this, but one generalization of interest is

$$\delta_{uj} = \underline{z}'_{uj}\underline{v}_u + \epsilon_{uj} + \underline{w}'_j\underline{q}_j$$

which allows random effects associated with the repeated measures positions. For example, if j indicates time then $\underline{w}'_j\underline{q}_j = q_j$ (scalar) indicates the presence of a ranom time effect q_j at time j.

Proc mixed for example will let you have more than one random statement with different variables playing the role of sub = which lets us have random effects in more than one direction. See the examples. It does not allow us however to have more than one repeated statement, which is a limitation.

Time Series/Cross-Sectional models involve exactly this notion of units (the cross sections) and repeated measures (time). These models can all be considered within the context of mixed models as described about, but the TSCS models and associated software have grown up somewhat independently.

1.4.1 Park's Model

Park's model combines serial correlation on each unit (with separate parameters for each unit) along with an additional error that can be cross correlated among units. This is good for equally spaced repeated measures over time, with the units/cross-sections treated as fixed.

$$Y_{uj} = \underline{x}'_{uj}\underline{\beta} + \delta_{uj}, j = 1...T \tag{4}$$

where

$$\delta_{uj} = \rho_u \delta_{u,j-1} + \epsilon_{uj},$$

$E(\epsilon_{uj}) = 0$, $E(\delta_{uj}) = 0$

$var(\epsilon_{uj}) = \phi_u^2$

$cov(\epsilon_{uj}, \epsilon_{wj}) = \phi_{uw}$. This allows correlation among different units at time j, which is often referred to as *contemporaneous correlation.*

$$\underline{\epsilon}_{.j} = \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ . \\ \epsilon_{Nj} \end{bmatrix}, \qquad Cov(\underline{\epsilon}_{.j}) = \Phi.$$

So, Φ is the covariance matrix among the ϵ terms at a fixed time t.

$var(\delta_{uj}) = \sigma_u^2 = \phi_u^2/(1-\rho_u^2)$.

$cov(\epsilon_{uj}, \delta_{u,j-1}) = 0$ (no correlation among current ϵ and previous δ.)

$cov(\epsilon_{uj}, \epsilon_{wk}) = 0, j \neq k$ (no correlation among ϵ's at different times).

Leads to $cov(\delta_{uj}, \delta_{wj}) = \sigma_{uw} = \phi_{uw}/(1-\rho_u\rho_w)$.

$$\Sigma_u = cov(\underline{Y}_u) = \sigma_u^2 P_{uu}, \qquad \Sigma_{uw} = cov(\underline{Y}_u, \underline{Y}_w) = \sigma_{uw} P_{uw},$$

where

$$P_{uw} = \begin{bmatrix} 1 & \rho_w & \rho_w^2 & ... & \rho_w^{T-1} \\ \rho_u & 1 & \rho_w & ... & \rho_w^{T-2} \\ \rho_u^2 & \rho_u & 1 & ... & \rho_w^{T-3} \\ . & . & . & . & . \\ \rho_u^{T-1} & \rho_u^{T-2} & \rho_u^{T-3} & ... & 1 \end{bmatrix}.$$

Here there are no random effects due to units so the units are treated as fixed. The $X_u\underline{\beta}$ part can be constructed so there may be separate coefficients in each unit (in which case X_u would involve dummy variables) or there could be common coefficients.

A special case of Park's model is where the ρ_u are all taken as 0. This often called a *Seemingly Unrelated Regressions* model though the term more generally can include models with the serial correlation.

The above yields a parametric model for Σ, and as in our earlier more general discussion we can estimate Σ to get either corrected standard errors for the least squares estimators or to get a generalized least squares estimator.

proc TSCSREG analyzes this model by using two-stage generalized least squares. Consider first a simple least squares fit and denote the residuals by r_{uj}. The correlations are often estimated via

$$\hat{\rho}_u = \sum_{j=2}^{T} r_{uj} r_{u,j-1} / \sum_{j=1}^{T-1} r_{uj}^2.$$

There are other estimates that are sometimes used.

These estimated correlations are then used to create transformed variables Y^*_{uj} and $\underline{x}^*_{uj}$ (as described in a single series with serial correlation) via $y^*_{u1} = (1-\hat{\rho}_u)^{1/2} y_{u1}$, and $\underline{x}^*_{u1} = (1-\hat{\rho}_u)^{1/2}\underline{x}_{u1}$,

$y^*_{uj} = y_{uj} - \hat{\rho}_u y_{u,j-1}$ and $\underline{x}^*_{uj} = \underline{x}_{uj} - \hat{\rho}_u \underline{x}_{u,j-1}$, $j = 2, \ldots, T$ leading to

$$\underline{Y}^* = X^* \underline{\beta} + \underline{\epsilon}^*,$$

If ρ's were known exactly then with $\underline{Y}^{*'} = [\underline{Y}^*_{.1}, \ldots, \underline{Y}^*_{.T}]$ with $\underline{Y}^{*'}_{.j} = [Y^*_{1j}, \ldots, Y^*_{uj}, \ldots, Y^*_{nj}]$ (so data is now grouped by time rather than "unit") then

$$cov(\underline{\epsilon}^*) = \Phi \otimes I = \begin{bmatrix} \Phi & 0 & . & 0 \\ 0 & \Phi & . & 0 \\ 0 & . & . & 0 \\ 0 & 0 & . & \Phi \end{bmatrix},$$

where $\otimes$ denotes Kronecker product.

We can now form a second estimator

$$\hat{\underline{\beta}}_{LS2} = (X^{*'} X^*)^{-1} X^{*'} \underline{Y}^*,$$

(which has removed the autocorrelation) and get residuals: $r^*_{uj} = Y^*_{uj} - x^{*'}_{uj}\hat{\underline{\beta}}_{LS2}$. The r^* residuals can be used to estimate Φ.

Proc TSCSREG uses $\hat{\phi}_{uw} = \sum_{j=1}^{T} r^*_{uj} r^*_{wj}/T$ (The SAS documentation indicates use of $T-p$, where $p = rank(X)$ but that is not what is done). It is better to use $\hat{\Phi}_{mod}$, where

$$\hat{\phi}_{uw(mod)} = \sum_{j=1}^{T} r^*_{uj} r^*_{wj}/(T-q).$$

q = number of coefficients entering into $E(Y_{uj})$ (assumed for simplicity to be the same for all uj.) This was Park's original suggestion. Also, there is no guarantee that the $\sum_j r^*_{uw} = 0$, so in general the values should be centered; TSCSREG does not do so.

Can now use the estimated ρ's and Φ to get $\hat{\Sigma}$, and get a generalized least squares estimator

$$\hat{\underline{\beta}}_{GLS} = (X'\hat{\Sigma}^{-1}X)^{-1} X'\hat{\Sigma}^{-1}\underline{Y},$$

with an estimated covariance of $\hat{\underline{\beta}}_{GLS}$ of

$$\hat{\Sigma}_{\hat{\beta}_{gls}} = (X'\hat{\Sigma}^{-1}X)^{-1}.$$

This describes how SAS proc TSCSREG does the analysis. More generally, the procedure could be iterated and there can be some modifications in how the "covariance parameters" are estimated. Note that the divisor in $\hat{\Phi}$ doesn't matter for $\hat{\underline{\beta}}_{GLS}$ but does for $\hat{\Sigma}_{\hat{\beta}_{GLS}}$

A problem with $\hat{\Sigma}_{\hat{\beta}_{gls}}$ is that the estimated variance/covariance matrix does not account for the uncertainty that goes in to estimating Σ. This can lead to possibly serious underestimation of the variance covariance matrix. Simulation results have shown that the estimated standard errors can be quite deficient. Bootstrapping for the estimation of the standard errors must be done in a way which accounts for the structure in the problem.

A Simple Bootstrap Approach to Park's model

We condition on the x values.

- r_{uj} = residuals from least squares (These should be centered so they have mean 0 over time. This will happen automatically with separate regressions per unit).

How to resample to mimic the δ_{uj}'s?

- Could use r^*'s (the residuals after the transformation to remove autocorrelation)
- We will instead use the residuals from the original fit to get

$$\hat{e}_{uj} = r_{uj} - \hat{\rho}_u r_{u,j-1}, j = 2, \ldots T.$$

Centered values: $\tilde{e}_{uj} = \hat{e}_{uj} - \hat{e}_{u.}$, where $\hat{e}_{u.} = \sum_{j=2}^{T} \hat{e}_{uj}/(T-1)$.

An alternate estimate of Φ is

$$\hat{\Phi}_{1mod} = \sum_{j=2}^{T} \underline{\tilde{e}}_{.j} \underline{\tilde{e}}'_{.j}/(T-1-q).$$

q = number of coefficients entering into $E(Y_{uj})$.

The distribution of the $\underline{\tilde{e}}_{.j}$ values can be viewed as a crude estimate of the distribution of the $\underline{\epsilon}_{.j}$'s. In resampling, will rescale so we are sampling from a set with covariance matrix $\hat{\Phi}_{1mod}$.

Generate B bootstrap samples. For sample b:

$$y_{buj} = \underline{x}'_{uj}\underline{\hat{\beta}}_{ls} + r_{bij}, j = 1, \ldots, T.$$

$j = 1$, $r_{bu1} = \left[\frac{T}{T-q}\right]^{1/2} r_{uk}$, where k is uniform over $1, \ldots T$

(at time 1 we sample from the least squares residuals directly.)

For $j = 2, \ldots, T$: $r_{buj} = \hat{\rho}_u r_{bu,j-1} + \left[\frac{T-1}{T-1-q}\right] \tilde{e}_{uk}$, where k is uniform over $2, \ldots T$.

Note that the k which is used is the same across all units. It is this which retains the contemporaneous part of the model.

1.4.2 Fuller-Battese

model.

The Fuller- Battese model merges random unit effects with cross unit correlation.

$$Y_{uj} = \underline{x}'_{uj}\underline{\beta} + \delta_{uj}, j = 1...T \tag{5}$$

where

$$\delta_{uj} = v_u + e_t + \epsilon_{uj},$$

where v_u, e_t and ϵ_{uj} all have mean 0, and are uncorrelated throughout with variances σ_v^2, σ_e^2 and σ_ϵ^2. This introduces a variance covariance form for Σ which depends on three parameters. v_u is an additive random unit effect, e_t is a random effect at time t which is common to all units and ϵ_{ut} captures additional noise. Note that this model is treating units as having additive random effects associated with them, which is the same as a random coefficients approach with random intercepts only.

TSCSREG fits this model by estimating the variance components (using what is known as the method of fitting constants which is a moment based method) and then does generalized weighted least squares. We can also fit it directly in proc Mixed.

1.4.3 Other models.

There are a number of other models that can be used. We can take any of the models for Σ_u we used earlier when we were assuming no among unit correlation and add in a contemporaneous correlation structure among

the ϵ_{uj} (over u at a fixed j). An important issue is having reasonable (statistically and computationally) ways to estimate the covariance parameters.

1.5 Analyzing the random coefficients linear mixed model using per "subject/unit" fits.

Suppose we have a random sample of units with the coefficients allowed to vary across units and the goal is inferences for the population coefficients. Can we make inferences for the population coefficients in a way that does not depend on us needing to model the covariance structure nor on requiring a common number of observations per unit or more strongly, a common design matrix per unit? The answer is yes. One can simply fit the linear model per subject and then analyze the fitted coefficient as the data across units. This result is shown in Buonaccorsi (2006) (as a special case of a more general formulation of the problem).

Random coefficients: for $u = 1$ to n, $\mathbf{B}_u$ are i.i.d. with mean $\boldsymbol{\beta}$ and covariance Σ_β. The random effects in our earlier formulation are $\mathbf{v}_u = \mathbf{B}_u - \boldsymbol{\beta}$.

$\mathbf{Y}'_u = (Y_{u1}, \ldots, Y_{um_u})$. $\mathbf{X}_u$ $(m_u \times p)$ design matrix associated with *uth* unit. Considered fixed.

Conditional Model: Given $\mathbf{B}_u = \mathbf{b}_u$,

$$\mathbf{Y}_u = X_u\mathbf{b}_u + \boldsymbol{\epsilon}_u, \quad E(\boldsymbol{\epsilon}_u|\mathbf{b}_u) = \underline{0}, Cov(\boldsymbol{\epsilon}_u|\mathbf{b}_u) = \Sigma_{eu}.$$

(Aside: Exactly what is meant by Σ_{eu} is a little subtle. Conditioning on $\mathbf{b}_u$, is different than conditioning on the unit. For here, view this as conditioning on what unit ends up in the *uth* position which determines $\mathbf{b}_u$. Note that Σ_{eu} might depend on $\mathbf{b}_u$, either explicitly or implicitly.)

Unconditional Linear Mixed Model:

$$\mathbf{Y}_u = X_u\boldsymbol{\beta} + \boldsymbol{\delta}_u,$$

where $\boldsymbol{\delta}_u = X_u(\mathbf{B}_u - \boldsymbol{\beta}) + \boldsymbol{\epsilon}_u$ has mean $\underline{0}$ and covariance $X_u\Sigma_\beta X'_u + E(\Sigma_{eu})$ (with Σ_β playing the role of what was $\mathbf{G}$ before.)

Suppose the objective is inferences for $\boldsymbol{\beta}$.

A simple approach using individual fits.

$\hat{\mathbf{b}}_u = (X'_uX_u)^{-1}X'_u\mathbf{Y}_u$ = least squares fit on unit u.

Conditionally:

$$E(\hat{\mathbf{b}}_u|\mathbf{b}_u) = \mathbf{b}_u, \;\; Cov(\hat{\mathbf{b}}_u|\mathbf{b}_u) = (X'_uX_u)^{-1}X'_u\Sigma_{eu}X_u(X'_uX_u)^{-1} = \Gamma_u.$$

Note that Γ_u is a conditional covariance and depends on Σ_{eu}, which is also conditional and may depend on random variables associated with the unit selected (including the $\mathbf{b}_u$ themselves.)

Unconditionally:

$$E(\hat{\mathbf{b}}_u) = \boldsymbol{\beta} \text{ and } Cov(\hat{b}_u) = \Sigma_\beta + (X'_uX_u)^{-1}X'_uE(\Sigma_{eu})X_u(X'_uX_u)^{-1}.$$

Define,

$$\hat{\boldsymbol{\beta}}_{mean} = \sum_u \hat{\mathbf{b}}_u/n \text{ and } S_b = \sum_u (\hat{\mathbf{b}}_u - \hat{\boldsymbol{\beta}}_{mean})(\hat{\mathbf{b}}_u - \hat{\boldsymbol{\beta}}_{mean})'/(n-1)$$

Result:

1. $E(\hat{\boldsymbol{\beta}}_{mean}) = \boldsymbol{\beta}$ (unbiasedness)

2. S_b/n is an unbiased estimator of $Cov(\hat{\boldsymbol{\beta}}_{mean})$.

MAIN POINT. This holds regardless of the covariance structure involved; that is regardless of the form of $Cov(\boldsymbol{\epsilon}_u|\mathbf{b}_u)$ or Σ_β. It also does not depend in any way on common design matrices or sample sizes across unit. It depends only on the $\hat{\mathbf{b}}_u$ being uncorrelated, each with expected value $\boldsymbol{\beta}$.

Can use this in usual way to get approximate confidence intervals and tests for the β's and linear combinations of them in the usual way. These analyses can be univariate or multivariate. As with all such problems normal based inferences will be suspect with small sample sizes and non-normality. Of course if the goal is to estimate Σ_β then something must be specified about the covariance structure.

1.6 Nonlinear mixed models.

The concepts above can be extended to handle non-linear models where the observations are uncorrelated. Things are a little more complicated because of the nonlinearity; in particular having a particular non-linear model for each unit does not mean the same non-linear model holds at the population level (viewed in terms of the model for $Y|X$ over a randomly selected unit). Here we describe some of the models used but do not provide a full treatment of the methods of analysis.

We use the same notation as in the linear mixed model; Y_{uj} is the jth observation on unit u with predictor vector $\underline{x}_{uj}$. With random effects, $\underline{v}_u$ is a set of random effects associated with unit u, with mean 0 and variance G.

There are two general approaches.

- **Unconditional or marginal model:** This is unconditional in the sense that it averages over any underlying random and looks at the marginal model of $Y_{uj}|X_{uj}$ over sampling of a unit to go into position u in the sample.

  $E(Y_{uj}|\underline{x}_{uj}) = \mu_{uj} = m(\underline{x}_{uj}, \underline{\beta})$, or in the generalized linear model, $\mu_{uj} = g^{-1}(\underline{x}'_{uj}\underline{\beta})$.

  In analyzing this we recognize that the components of $\underline{Y}_u$ can be correlated, with $Cov(Y_u) = \Sigma_u$.

  This model is often analyzed through the use of GEE (generalized estimating equations) which will use some working assumption about Σ_u but then get the estimated covariance of $\hat{\underline{\beta}}$ in a way that is robust to the specification of Σ_u. Often the working assumption about Σ_u is independence so the usual analysis is run (this may be a general non-linear model or logistic regression, etc.) Then a robust estimate of $Cov(\hat{\underline{\beta}})$ is obtained. Proc Genmod in SAS analyzes such a model and the documentation there provides a succinct description of the methodology, which has much in common with the approach in linear models.

- **Conditional model on the unit level. Nonlinear mixed model**: This specifies a model given the random effects associated with the uth selected unit.

  $E(Y_{uj}|\underline{x}_{uj}, \underline{v}_u) = \mu_{cuj} = m(\underline{x}'_{uj}, \underline{z}_{uj}, \underline{v}_u, \underline{\beta})$.

  For a **generalized linear mixed model** $\mu_{cuj} = g^{-1}(\underline{x}'_{uj}\underline{\beta} + \underline{z}'_{uj}\underline{v}_u)$ or $g(\mu_{cuj}) = \underline{x}'_{uj}\underline{\beta} + \underline{z}'_{uj}\underline{v}_u$, where g is the link function.

  In this case the full distribution of the data is $\prod_u f(\underline{Y}_u|X_u, \underline{\theta}, \underline{v}_u) f_v(\underline{v}_u|\underline{\xi})$ where f_v is the density or mass function of the v_u involving parameters $\underline{\psi}$. The term inside the product gives the marginal distribution of $\underline{Y}_u$. Estimation and inferences are based on this likelihood. Proc Nlmixed will fit such models.

Notice that if the conditional model is linear, that is $E(Y_{uj}|\underline{x}_{uj}, \underline{v}_u) = \underline{x}'_{uj}\underline{\beta} + \underline{z}'_{uj}\underline{v}_u$, then the marginal model is also linear, i.e., $E(Y_{uj}|\underline{x}_{uj}) = \underline{x}'_{uj}\underline{\beta}$. This is a nice feature of the linear models (which remember includes models that are polynomial in the predictors.) When the conditional model is non-linear, the

functional form of the mean/regression model in the conditional model *does not carry over to the marginal model!* To a first order approximation however the marginal model is of the same for with coefficients equal to average coefficients over units. But the accuracy of this for the marginal model depends on the relative variance of the coefficients over units.

1.7 References

- CH - Crowder and Hand, "Analysis of Repeated Measures" (Chapters 5 and 6 in particular)
- Hand and Crowder, "Practical Longitudinal Data Analysis", Chapman and Hall.
- Diggle, Liang and Zeger. "Analysis of Longitudinal Data", 1994. Oxford.
- P. J. Diggle, P. Heagerty, K.-Y. Liang, and S. L. Zeger (1994). " Analysis of Longitudinal Data." 2nd Edition, Oxford Science Publications.
- J. Ware - "Linear Models for the Analysis of Longitudinal Studies", American Statistician, 1985, 95-101;
- "Random Coefficient Models", Lonford. Oxford Press.
- "Mixed Effects Models in S and S-Plus", Pinheiro and Bates. Springer.
- Fitzmaurice GM, Laird NM and Ware JH. (2004). Applied Longitudinal Analysis. New York: John Wiley and Sons.
- SAS System for Mixed Models. Littell, Milliken, Stroup and Wolfinger. Published by SAS.
- Judge et. al. "Theory and Practice of Econometrics". (Ch. 13)
- Griffiths et al. "Learning and Practic of Econometrics",Chapter 17.
- Generalized, Linear and Mixed Models. McCulloch and Searle, John Wiley.
- Mixed Models. Eugene Demidenko.
- Christiansen. "Plane Answers to Complex Questions". Chapter 12.
- Hocking. The Analysis of Linear Models.
- Davidian and Giltian. Nonlinear Models for Repeated Measures Data.
- McCulloch, Searle and Neuhaus. Generalized, Linear and Mixed Models.
- Littell,Milliken, Stroup and Wolfinger. SAS System for Mixed Models, 2nd Edition. SAS Publishing.