

# Estimation without $E(\varepsilon\varepsilon'|X) = \sigma^2I$

C. Flinn

November 29, 2000

## 1. GLS and OLS with Known Covariance Matrix

The following results apply to the linear regression model

$$y = X\beta + \varepsilon,$$

where  $X$  is of dimension  $(n \times k)$ ,  $\varepsilon$  is a (unknown)  $(n \times 1)$  vector of disturbances, and  $\beta$  is a (unknown)  $(k \times 1)$  parameter vector. We assume that  $n \gg k$ , and that  $\rho(X) = k$ . This implies that  $\rho(X'X) = k$  as well.

Throughout we assume that the “classical” first conditional moment assumptions apply, namely

$$E(\varepsilon_i|X) = 0 \quad \forall i.$$

However, we relax the second conditional moment condition to be

$$E(\varepsilon\varepsilon'|X) = \sigma^2\Psi.$$

**Proposition 1.1.** *The BLUE of  $\beta$  when  $E(\varepsilon\varepsilon'|X) = \sigma^2\Psi$  is given by  $\tilde{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$ .*

**Proof.** Given that  $\Psi$  is finite, symmetric and nonsingular, it possesses an inverse  $\Psi^{-1}$ . By the symmetry of  $\Psi$ , there exists a matrix  $L$  such that  $L'L = \Psi^{-1}$ . Then take the original regression equation and multiply both sides through by  $L$ , so

$$\begin{aligned} Ly &= LX\beta + L\varepsilon \\ \Rightarrow \tilde{y} &= \tilde{X}\beta + \tilde{\varepsilon}. \end{aligned}$$

Note that  $E(\tilde{\varepsilon}|\tilde{X}) = E(L\varepsilon|LX) = 0$ , so that mean independence continues to hold in the “transformed” regression function. Most importantly, note that

$$\begin{aligned} E(\tilde{\varepsilon}\tilde{\varepsilon}'|\tilde{X}) &= LE(\varepsilon\varepsilon'|\tilde{X})L' \\ &= \sigma^2 L\Psi L'. \end{aligned}$$

But  $L'(L\Psi L') = \Psi^{-1}\Psi L' = L' \Rightarrow L\Psi L' = I \Rightarrow E(\tilde{\varepsilon}\tilde{\varepsilon}'|\tilde{X}) = \sigma^2 I$ . Then the equation  $Ly = LX\beta + L\varepsilon$  satisfies the full ideal conditions on the first two conditional moments, thus implying that the “OLS” estimator associated with this equation is BLUE. This “OLS” estimator is just  $\tilde{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$ .  $\nexists$

Then by applying the transformation  $L$  to the regression function, the resulting estimation problem is just one of OLS. The procedure works because linear transformations do not alter the relationship between  $y$  and  $X$ . In terms of estimation, when one knows the matrix  $\Psi$  we can directly form the estimator  $\tilde{\beta}$  by inverting  $\Psi$ . As some of you found out, this can be problematic since this matrix is of dimension  $(n \times n)$ . Instead of this “brute force” procedure, it is often more tractable to first perform the transformation of the data and then proceed to perform OLS on the transformed data. Consider the following example.

Example 1.2. *Let*

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

with  $E(\varepsilon|x) = 0$  and  $V(\varepsilon_i|x_i) = \sigma^2 x_i^2$ ,  $E(\varepsilon_i \varepsilon_j|x) = 0$  all  $i \neq j$ . Then we

$$E(\varepsilon\varepsilon'|x) = \sigma^2 \begin{bmatrix} x_1^2 & 0 & \dots & 0 \\ 0 & x_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & x_n^2 \end{bmatrix},$$

which implies that

$$\Psi^{-1} = \begin{bmatrix} x_1^{-2} & 0 & \dots & 0 \\ 0 & x_2^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & x_n^{-2} \end{bmatrix},$$

which implies

$$L = \begin{bmatrix} x_1^{-1} & 0 & \dots & 0 \\ 0 & x_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & x_n^{-1} \end{bmatrix}.$$

The new dependent variable  $\tilde{y} = Ly = [y_1/x_1 \dots y_n/x_n]'$ . The new matrix of regressors is

$$\begin{aligned}\tilde{X} &\equiv LX = L \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \\ &= \begin{bmatrix} 1/x_1 & 1 \\ 1/x_2 & 1 \\ \vdots & \vdots \\ 1/x_n & 1 \end{bmatrix}.\end{aligned}$$

The new error vector has mean 0 and covariance matrix  $\sigma^2 I$ . Then the GLS estimator  $\tilde{\beta}$  is equal to  $(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$ .

Since  $\tilde{\beta}$  is the OLS estimator for the transformed data, it follows that

Proposition 1.3.  $V(\tilde{\beta}) = \sigma^2(X'\Psi^{-1}X)^{-1}$ .

Proof. The OLS estimator from the transformed data has covariance matrix

$$\begin{aligned}\sigma^2((LX)'(LX))^{-1} &= \sigma^2(X'L' LX)^{-1} \\ &= \sigma^2(X'\Psi^{-1}X)^{-1}.\text{✎}\end{aligned}$$

We may not know  $\sigma^2$  a priori, though we continue to assume that we do know  $\Psi$ . We have the following “good” estimator of  $\sigma^2$  in this case.

Proposition 1.4. An unbiased estimator of  $\sigma^2$  is

$$\tilde{\sigma}^2 = \frac{\tilde{a}'\Psi^{-1}\tilde{a}}{n - k}$$

where  $\tilde{a} \equiv y - X\tilde{\beta}$ .

Proof. Since transformed equation satisfies full ideal conditions, the desired “OLS” estimator of  $\sigma^2$  is

$$\begin{aligned}&(n - k)^{-1}(Ly - LX\tilde{\beta})'(Ly - LX\tilde{\beta}) \\ &= (n - k)^{-1}(L(y - X\tilde{\beta}))'(L(y - X\tilde{\beta})) \\ &= (n - k)^{-1}(y - X\tilde{\beta})'L'L(y - X\tilde{\beta}) \\ &= (n - k)^{-1}(y - X\tilde{\beta})'\Psi^{-1}(y - X\tilde{\beta}).\text{✎}\end{aligned}$$

All the usual hypothesis tests can be applied to the “transformed” equations. For example, if  $\varepsilon$  is normally distributed, the estimator  $\tilde{\beta}$  will be normally distributed as well. For tests of linear restrictions on the  $\beta$  vector, small sample  $t$  and  $F$  tests will be appropriate.

Before moving on to more general cases, we look at conditions required for consistency of the OLS estimator under conditions of heteroskedasticity. Now

$$\begin{aligned}\text{plim } \hat{\beta} &= \text{plim}(X'X)^{-1}X'y \\ &= \beta + \text{plim}(X'X)^{-1}X'\varepsilon \\ &= \beta + \text{plim}\left(\frac{X'X}{n}\right)^{-1} \text{plim}\left(\frac{X'\varepsilon}{n}\right).\end{aligned}$$

But since  $\left(\frac{X'X}{n}\right)^{-1} \rightarrow Q^{-1}$ , a finite nonsingular matrix, and since  $E\left(\frac{X'\varepsilon}{n}\right) = 0$  for all  $n$ , then

$$V\left(\frac{X'\varepsilon}{n}\right) = \frac{\sigma^2 X'\Psi X}{n^2}.$$

Now if

$$\lim_{n \rightarrow \infty} \frac{X'\Psi X}{n} < \infty,$$

then

$$\lim_{n \rightarrow \infty} \frac{X'\Psi X}{n^2} = 0.$$

Since the mean of  $\frac{X'\varepsilon}{n}$  is 0 for all  $n$  and since the limit of the variance of the random variable vanishes asymptotically,  $\text{plim } \hat{\beta} = \beta$ .

If we don't assume that  $\varepsilon$  is normally distributed, in large samples the covariance matrix of  $\hat{\beta}$  is

$$\frac{\sigma^2}{n} \lim\left(\frac{X'X}{n}\right)^{-1} \lim \frac{X'\Psi X}{n} \lim\left(\frac{X'X}{n}\right)^{-1}.$$

## 2. Feasible GLS

The assumption of a known  $\Psi$  is clearly unrealistic in most situations. Since the covariance matrix of the errors is  $n \times n$ , but is restricted to be positive definite and symmetric, there are  $n \times (n + 1)/2$  free (but subject to restrictions such that  $\Psi$  is positive definite) elements. This is clearly more than can be estimated with  $n$  pieces of information. All is not lost however.

Recall that if a known function  $g$  is sufficiently smooth, and if  $\hat{\theta}$  is a consistent estimator for some parameter  $\theta$ , then  $\text{plim } g(\hat{\theta}) = g(\theta)$ . Now the GLS estimator is

$$\tilde{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y.$$

We can think of this estimator as a function of  $\Psi$ . If we don't know  $\Psi$ , then the natural second best is to substitute an estimator with desirable properties. Let our estimator be denoted  $\hat{\Psi}$ . For asymptotic results to go through, we will generally require that our estimator of  $\Psi$  be consistent. But how can that be if a new set of parameters is added every time a new observation is added? One must obviously resort to imposing some restrictions on the form of  $\Psi$ .

*Definition 2.1. If  $\Psi$  depends on a fixed, finite number of parameters  $\theta_1, \dots, \theta_p$  and if we have consistent estimators of each of these parameters,  $\hat{\theta}_1, \dots, \hat{\theta}_p$ , and if the  $\Psi(\theta)$  is sufficiently smooth, then  $\hat{\Psi} = \Psi(\hat{\theta})$  is a consistent estimator of  $\Psi$ .*

Now consider the feasible GLS estimator, which is essentially

*Definition 2.2. The Feasible GLS (FGLS) estimator of  $\beta$  is given by*

$$\tilde{\beta}(\hat{\theta}) = (X'\Psi(\hat{\theta})^{-1}X)^{-1}X'\Psi(\hat{\theta})^{-1}y.$$

Now in practice we usually solve for the FGLS estimator in a number of steps. First we require consistent estimates of the parameters characterizing  $\Psi$ . These are typically obtained by using analog type estimators which employ moments of OLS residuals. To get the OLS residuals obviously requires that we first find the OLS estimate of  $\beta$ ,  $\hat{\beta} = (X'X)^{-1}X'y$ . After we obtain the required estimates  $\hat{\theta}$ , we form  $\hat{\Psi} = \Psi(\hat{\theta})$ , and then find the FGLS estimates. Consider the following example.

*Example 2.3. Let*

$$E(\varepsilon\varepsilon'|X) = \sigma^2 \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha \\ \alpha & \dots & \alpha & 1 \end{bmatrix},$$

where  $|\alpha| < 1$ . Now this covariance structure implies that the variance of any given disturbance is  $\sigma^2$ , and the covariance is between any two observations  $i$  and  $j$  is

$\alpha\sigma^2$  for all  $i \neq j$  [note that the correlation coefficient is simply  $\alpha$ ]. Now we know that OLS is consistent in the general  $\Psi$  case, and that therefore asymptotically OLS residuals will converge to their corresponding disturbances. First estimate  $\hat{\beta} = (X'X)^{-1}X'y$ . Now define the OLS residual vector by  $a \equiv y - X\hat{\beta}$ . Now define the estimator

$$\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n a_i^2.$$

This is the estimated sample variance. By the law of large numbers, its probability limit is the corresponding sample moment.

By the same token, define the sample covariance between all pairs of disturbances in the sample, which is

$$\hat{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^n a_i a_j.$$

Given the existence of the appropriate population moments, the probability limit of  $\hat{c}$  will be  $c = \sigma^2\alpha$ . Then a consistent estimator of  $\alpha$  is given by

$$\hat{\alpha} = \hat{c}/\hat{\sigma}^2.$$

Armed with these estimates, we form

$$\hat{\Psi} = \begin{bmatrix} 1 & \hat{\alpha} & \dots & \hat{\alpha} \\ \hat{\alpha} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\alpha} \\ \hat{\alpha} & \dots & \hat{\alpha} & 1 \end{bmatrix}.$$

We then compute the FGLS estimator using this  $\hat{\Psi}$ .

Before saying something about the asymptotic distribution of the FGLS estimator, we note that in finite samples the estimator will differ from a m.l. estimator which estimates the parameters  $\beta$  and  $\theta$  simultaneously. Note that the FGLS procedure first obtains one estimator for  $\beta$ ,  $\hat{\beta}$ , upon which estimates of  $\Psi$  are based. Using these estimates, another estimator of  $\beta$  is obtained,  $\tilde{\beta}(\hat{\theta}) = \tilde{\beta}(\theta(\hat{\beta}))$ . The estimation procedure does not insure that  $\hat{\beta} = \tilde{\beta}$ . Essentially, the maximum likelihood procedure will not form two separate estimators of  $\beta$ , and thus will not suffer from this schizophrenic behavior. Differences between GLS and FGLS, and between both and m.l., will generally disappear asymptotically.

Proposition 2.4. *A sufficient condition for the FGLS estimator to be consistent is that*

$$\text{plim} \frac{X' \hat{\Psi}^{-1} X}{n} \text{ is finite and nonsingular}$$

and

$$\text{plim} \frac{X' \hat{\Psi}^{-1} \varepsilon}{n} = 0.$$

Proposition 2.5. *A sufficient condition for the GLS and FGLS estimators to have the same asymptotic distribution is*

$$\text{plim} \frac{X'(\hat{\Psi}^{-1} - \Psi^{-1})X}{n} = 0$$

and

$$\text{plim} \frac{X'(\hat{\Psi}^{-1} - \Psi^{-1})\varepsilon}{n} = 0.$$

The proofs of both of these results are straightforward, being mainly definitional. While there exist some counterexamples, in general if you have access to a consistent estimator of the matrix  $\Psi$ , the same asymptotic tests available using  $\Psi$  can be conducted using the consistent estimator of  $\Psi$ .