Econometrics notes round 1.

**Chapter 7: GLS and related topics**

So far, we assume exogeneity and homoskedasticity. We need new estimators (for unbiasedness, consistency, and/or efficiency) if these assumptions do not hold (however, we need to note that, if these assumptions are violated, we should always think about the possibility of misspecification of our models, and we use specific estimators which we are about to introduce as the last resort).

Suppose we have the following (linear) model:

(7.01)

where is a positive definite *n* x *n* matrix (because is a symmetric matrix). is equal toif we assume exogeneity and homoskedasticity. However, this may not be true, e.g., usually, for time series data, we have errors autocorrelated and the correlations are highest for observations which are close.

Under this circumstance, OLS is still unbiased.

However, the variance-covariance matrix for the parameter is incorrect, thus is biased (usually downwards) even asymptotically.Therefore, interval estimation and hypothesis testing for the parameter are unreliable.

A mitigation is to use heteroskedasticiy-consistent and autocorrelation-consistent variance-covariance matrix estimators for the OLS estimator to eliminate the asymptotic bias (but not completely the small sample bias) of . This is sometimes called a robust variance estimator.

**7.2 the GLS estimator**

We can obtain an **efficient** estimator for (7.01) by transforming the model to satisfy the requirement by the Gauss Markov Theorem.

It can be proven that we can always find a non-singular *n* x *n* triangular matrix (i.e., this is by definition as is a positive definite matrix) so that:

(7.02)

We can pre-multiply (7.01) by and we have:

(7.03)

We can compute the OLS estimate for (7.03) and we call the OLS estimate for (7.03) the **GLS estimator**:

(7.04)

It is obvious that the variance-covariance of the error term for (7.03) is the identify matrix:

Also, as we know that, the variance-covariance for the OLS estimate is:

(3.28)

Thus, we can compute the variance-covariance matrix of the GLS estimator by replacing by and by 1 in (3.28), and we have:

(7.05)

Thus, if we know we can use the GLS estimator which drops the assumption of exogeneity and homoskedasticity.

Computationally the GLS estimator of can be calculated by minimizing the GLS criterion function:

(7.06)

Which is simply the **SSR** from the transformed regression (e.g., (7.03)). This can be thought of as a generalization of the **SSR** function in which the squares and cross products of the residuals from the original regression (7.01) are weighted by the inverse of the matrix .

**Efficiency of the GLS estimator**

It can be proven that GLS estimator, e.g., , in most cases is more efficient, and is never less efficient than the OLS estimator. Intuitively, when is equal to, this is because reduced to. If is not equal to, is more efficient than as take advantages of by discriminating observations which are more / less informative.

**7.3 Computing GLS estimates**

However, we do not know , and it is impossible to estimate the elements in because there are conceptually different elements in and we only have observations. Therefore, practical procedures for GLS estimation require us to assume quite a lot about the structure of the covariance matrix and its inverse. e.g., we have to assume that the non-sphericalness of the disturbance takes a specific form within one of the general categories of heteroskedasticity or autocorrelated disturbances. This changes the problem to one of assuming an appropriate specific form, estimating a small number of parameters (usually only one) that characterize that specific form, and then using these estimates to produce the required estimate of .

Also, we know that we can estimate the parameter without knowing the error variance. Therefore, suppose , we can estimate the parameter without knowing the value of – we only need to know the structure of the covariance matrix . This is because:

This suggests that we can estimate without knowing. We only need to know which describes the structure of the covariance matrix . However, we still do need to know the full content of covariance matrix if we want to calculate the variance matrix of the parameter:

where is the usual OLS estimate of the error variance for the transformed regression in (7.03).

**Weighted LS**

It is easy to obtain GLS estimates when the variance matrix for the error term to be heteroskedastic but serial uncorrelated. This indicates that is diagonal. We denote as the element in the matrix, thus, we denote as the element in the matrix. Accordingly, as we know that

(7.02)

Thus, is a diagonal matrix with the element as . Therefore, for a typical regression like (7.03), we have:

(7.12)

We can estimate this regression using OLS as it is obvious that the error variance is 1.

Thus, we say that (7.12) is a regression model estimated using **weight least square** (**WLS**) and the weight for observation *t* is . Observations with large error variance are given lower weights. In practice, if is not diagonal but with known structure, e.g., , (7.12) is still valid, except that we interpret as the element of and the variance of error term is now instead of 1.

There are many ways to determine the weights to be used in WLS. For example, we may assume that the diagonal elements of the variance matrix of the error term (e.g., ) to depend on some other variables which we observe, say, (e.g., in some economic studies, it could be population or national income). Another example is where the data we observe are obtained by grouping on different number of individuals. It is possible that ungrouped data have constant a variance, but each observation is an average of units, where varies. Thus, the variance of is proportional to . Thus, in this case, plays the role of in (7.12).

**Generalized NLS**

GLS is also applicable to nonlinear regression models. if the vector of regression functions were instead of , we could obtain generalized nonlinear least squares, or **GNLS** estimates by minimizing the following criterion function:

(7.13)

(7.13) is similar to (7.06):

If we differentiate (7.13) by and divide by -2, we have:

(7.14)

where is the matrix of derivatives of with respect to .

Finding estimates that solve (7.14) requires some sort of nonlinear minimization procedure.

**7.4 Feasible Generalized Least Squares**

In practice, is often not known even up to a scalar factor. Thus, we cannot compute GLS estimates. However, in many cases, it is reasonable to suppose that ordepends in a known way on a vector of unknown parameter . If so, we can estimate consistently to obtain . Also, because , once we estimate , we can estimate ), and we can compute GLS estimates conditional on ). This procedure is called **Feasible GLS** because it is feasible in many cases when ordinary GLS is not feasible.

A simple example:

(7.18)

where andare respectively a *k*-vector and an *l*-vector of unknown parameters. and are row vectors containing variables for observation *t*. In (7.18) we assume that some or all of elements in belong to **.** The function ofis a **skedastic** function. As a regression function determines the conditional mean of a r.v, a skedastic function determines the conditional variance of a r.v.

To obtain consistent estimate of , we first calculate consistent estimate of the error term in (7.18). e.g., we can calculate the OLS estimate first,and subsequently the OLS residual, e.g., . We can then run the auxiliary linear regression:

So that we can find the OLS estimate of . We then use the estimate of of to calculate the weights, e.g., for all *t*.

Finally, we can calculate the FGLS estimate of by using OLS estimator for the regression (7.12) by replacing the unknown with the estimated .

In this example, we have assumed that the error terms are heteroskedastic but serial uncorrelated.

**Why FGLS works**

Under suitable regularity conditions, it can be shown that is consistent and asymptotically equivalent to . Whether or not FGLS is a desirable estimation method in practice depends on how good an estimate of can be obtained.

Thus, we somehow take into account the nonsphericalness of the disturbance. As a result, with FGLS, is asymptotically unbiased. However, note that FGLS is biased and nonlinear.

It is possible to iterate an FGLS procedure. E.g., the estimator can be used compute a new set of residuals which can be used to calculate a second-round estimate of , which can be used to calculate a second-round estimate of and so on.

**7.5 Heteroskedasticity**

There are two situations where the error terms are heteroskedastic but not serially correlated:

1. The form of Heteroskedasticity is completely unknown.
2. The skedastic function is known except for the values of some parameters which can be estimated consistently.

If we know the error variance or at least , we can have efficient estimates using WLS. If we know the form of the Heteroskedasticity but the skedastic function depends on unknown parameters, we can then use feasible WLS and still achieve asymptotic efficiency.

However, before we apply feasible WLS, it is advisable to perform a specification test of the null hypothesis that the error terms are homoscedastic against whatever heteroskedastic alternatives may seem reasonable.

**Testing for Heteroskedasticity**

A reasonable general model of conditional Heteroskedasticity is:

(7.24)

where is a nonlinear function which only takes positive values, is a 1 x *r* vector of observations on exogenous or predetermined variables that belong to the information set which we believe determines . is a scalar parameter. is a *r*-vector of parameter. Under the null hypothesis of , the function reduces to , a constant, and there is no Heteroskedasticity. We can rewrite (7.24) as:

(7.25)

where is the difference between and the conditional expectation of .

In practice, a plausible specification of the skedastic function, e.g., ,is:

However, the test for Heteroskedasticity does **not** really depends on the function form of . It can be proven that no matter what function forms we choose, (7.25) can always be simplified as:

where and are redefined parameters (e.g., they can be related to the function form of ). In practice, we do not observe , and thus we replace with its estimates, e.g., , and the regression model becomes:

(7.28)

Thus, the null hypothesis of in (7.25) is equivalent to the null hypothesis of in (7.28).

The test of Heteroskedasticity in (7.28) requires us to choose . If economic theory does not tell us how to choose , there is no simple rule to choose it. Ideally, we would prefer adding variables which effectively explains . One approach is to choose functions of some original regressors, e.g., some suggest including squares and cross products of all the regressors. However, this causes to be very large and the test would have poor finite-sample properties unless we have a very large sample.

**7.6 Autoregressive and Moving-Average Process**

In practice, the error terms of the regression model may be autocorrelated. This normally indicates a misspecification of the model itself, but in some cases, it may be reasonable to model the error terms as a stochastic process.

Thus, if we believe there could be serial correlations for the error terms, we need to conduct the corresponding specification test. If we find autocorrelation for the error terms, we estimate a model which accounts for it.

**The AR(1) process**

The AR(1) process is most commonly seen in the error term. e.g.,

(7.29)

The condition of is called a stationary condition because is necessary for (7.29) to be stationary.

is called **covariance stationary** (or **wide sense stationary**) if:

1. the unconditional mean, e.g., and unconditional variance exist and they are independent of *t*.
2. the covariance, e.g., also exists and is independent of *t*, given any *j*.

We can rewrite (7.29) recursively by substituting the lagged error terms and we have (note that although we only observe the data for a limited period the data series exist for an infinite time):

Thus, we have the variance of as:

(7.31)

Thus, we see that exists and approaches , and is independent of *t* as required by the stationary condition. Note that (7.31) can only hold when (e.g., (7.31) can only converges when which is required by the formula which sums values from an equal ratio series). Without , the variance of the error terms would increase without limit as sample sizes increase. Therefore, the condition of is called a stationary condition which is necessary for to be stationary.

Also, we see that the covariance is:

As we assume predeterminedness in (7.29), , and the equation above becomes:

Therefore, exists and is independent of *t* for any *j* (given , as we need to exist in the first place).

We can then calculate the correlation of by dividing by the square root of the product of the variance of , that is, by . It is obvious that we find the correlation of is just .

We can take more steps (omitted here) to calculate the variance-covariance matrix of the error term vector as:

(7.32)

We note that the first component of (7.32) is just the variance of , as shown in (7.31). We denote the matrix component in (7.32) as the **autocorrelation matrix**, e.g., .

**High-order autoregressive process**

A more general stochastic process is the order autoregressive process, e.g., AR(p) process:

(7.33)

(7.33) can be rewritten as:

(7.34)

where is the lag operator. In order to make (7.33) stationary, we need to have restrictions on the parameters including (just like the restriction of for (7.31)).

Usually for convenience we would like to rewrite (7.33) in a more compact form. Suppose that, for an arbitrary we define a function of as:

(7.35)

Thus, we can then rewrite (7.33) using the function of as:

However, the disadvantage of this kind of compact expression is that it is inconvenient for us to describe the restrictions on . Thus, we use an alternative way to describe those restrictions we have on . e.g., we say that in order to make (7.33) stationary, for the following equation:

Or equivalently,

The roots must be lie outside the **unit circle** (e.g., greater than 1 in absolute value). The roots of the equation above are about (expressed in ). Thus, when we say the roots of the equation must be greater than 1 in absolute value, what we have done is to put restrictions on A special case is in (7.29) where we have:

Thus, the unit root is . In order to make (7.29) stationary, we need , which suggests .

**Moving-Average process**

An alternative type of stochastic process is moving average (**MA**) process. The simplest example is MA(1):

(7.37)

It can be proven that under the MA(1) process, and both exist and independent of *t*:

(7.37) also indicates that (thus the correlation as well) is zero for any .

Note that since , . (this is because and are independent, thus we have , and thus we have according to law of iteration).

We can calculate the covariance matrix for the vector as:

(7.38)

(7.38) suggests there is no correlation between error terms which are more than one period apart.

Just like the AR(p) process, we can write the MA(p) process as:

Or alternatively,

where is a polynomial in the lag operator.

**7.7 Testing for serial correlation**

Suppose we have the following model:

We can rewrite them as:

(7.41)

Thus, if we want to test for autocorrelation in the error term, we just need to estimate (7.41) using NLS and calculate the *t* statistic for . Note that (7.41) has a lagged dependent variable as the regressor, thus the *t* statistic does not follow a student’s t distribution in finite sample (even if the error terms are normally distributed) but only asymptotically.

However, usually we conduct the test for serial correlation not by testing (7.41). instead, we conduct tests based on the Gauss-Newton regression for some advantages (e.g., being valid under weak assumptions such as predeterminedness, lower computational cost, and also available for higher-order autoregressive process and even moving average process).

If we wish to test against an MA(q) process, we can proceed exactly as if we were testing against an AR(q) process. This is because AR process is locally equivalent to a MA process of the same order.

**DW test**

**Monte Carlo tests for serial correlation**

**Bootstrap test for serial correlation**

**Heteroskedasticity-Robust tests**

The tests for serial correlation what we have discussed so far are all based on the assumption of homoscedastic error terms. If this assumption is violated, the asymptotic distributions of all these test statistics differ from whatever distributions they are supposed to follow asymptotically. It is easy to twist these tests to make them Heteroskedasticity-Robust. However, the resulted Heteroskedasticity-Robust test is asymptotically valid and is not exact in finite samples.