Heteroscedasticity-Consistent (HC) Standard Errors and SAS

Ordinary Least Squares (OLS) estimates for a linear model are easy to compute using SAS and PROC REG, but any violation of the OLS assumption that the error terms are independent and identically distributed (i.i.d.) can create problems with using the OLS results. While the estimated coefficients are unbiased and asymptotically consistent under numerous conditions, the standard errors and t-statistics for the coefficients may biased and asymptotically inconsistent (i.e. not robust) under the same conditions.

W

Heteroscedasticity, which refers to variation in the variance of the model's error term, is one condition that may need to be controlled for or corrected by modifying the OLS formulas. SAS can produce HC covariance matrices in a variety of its regression-based procedures. PROC REG has an ACOV option that produces an asymptotically-consistent covariance matrix, while PROC SURVEYREG uses such a matrix as its default. For the greatest flexibility, PROC MODEL can be used to implement a Generalized Method of Moments (GMM) procedure that will yield not only a HC covariance matrix but a more general heteroscedasticity and autoregressive consistent (HAC) covariance matrix (also known as the Newey-West correction). Another option is to use PROC IML to explicitly calculate a HC covariance matrix.

Three different ways of using SAS to produce HC covariance matrices are demonstrated in links below, using a CAPM motivated application to the Fama-French size (SMB-- Small Minus Big) and value (HML-- High Minus Low) factors.

Overview of Heteroscedasticity-Consistent (HC) Covariance Matrices

When heteroscedasticity is present, OLS formulas do not properly calculate standard errors for the estimated coefficients of a regression equation. Ignoring heteroscedasticity can bias the estimated standard errors and associated tests or *p* values, especially when it is severe or related to the explanatory variables. (The direction of the bias depends on many factors, but more often than not the OLS standard errors are too small.)

Sometimes the form of the heteroscedasticity is obvious and it can be corrected using weighted least squares (i.e., it can be modeled). It is more common, however, that the form of the heteroscedasticity is unknown or impossible to adequately model. In this case, a heteroscedasticity-consistent (HC) covariance matrix can be used by using what is called the White correction by econometricians, or equivalently, using what is called the Huber sandwich estimator by statisticians.

Econometric Details

The well-know formula for OLS estimation is:

ßols = inv(X'X)*(X'y)

where X is the matrix of explanatory variables, y is the dependent variable, and inv() is the inverse function. The variance of ßols is equal to the variance of inv(X'X)*(X'e), where e is a vector of the model's random term (y-Xß). Hence:

Var(ßols) = inv(X'X)*(X'(e e') X)*inv(X'X)

If e is i.i.d. then E [(ee')] = var(e)*I , such that this matrix has a constant along its diagonal. In this case, the variance of ßols reduces to var(e)*inv(X'X). But if e is not identically distributed, such that var(e) is not constant, the more complicated formula should be used.

The equivalent Generalized Method of Moments (GMM) formula that accounts for heteroskedasticity is

var(ß) = inv(D)*(S)*inv(D)

where D=(X'X) and S=X'V'X in the linear case. Assuming that e is independent (i.e. shows no serial correlation), V is a diagonal matrix of E[e*e] elements. One way to compute S is to multiply each row of X by the corresponding row of e (based the OLS residuals) and then compute the cross product of this adjusted X matrix.

As long as the error terms are independent, the resulting HC covariance matrix is asymptotically correct in the sense that as the sample size increases, it converges to a consistent and unbiased estimate of the variance of the OLS coefficients. MacKinnon and White (1985) developed closely related, but alternative estimates that reduce the effects of outliers and supposedly have better finite sample properties. Newey and West (1987) show how to control for autocorrelation.

CAPM Application to Fama-French Factors

Because the Fama-french factors are effectively stock portfolios, a CAPM equation can be applied. CAPM theory implies that

E [Rj – Rf] = ß E [Rm- Rf]

Here Rj is the return on either the SMB or HML portfolios, Rm is the market return, Rf is the risk free rate, and ß = cov(Rj,Rm-Rf) / var(Rm-Rf)

An empirical application of this equation is

Rj(t) = a + b * (Rm(t)-Rf(t)) + c * Rf(t) + e(t)

Under the CAPM null hypothesis, a=0 and c=1 is expected. The c=1 restriction is most easily tested by subtracting Rf(t) from the left hand side and testing c'=0 in the equation

Rj(t)-Rf(t) = a + b * (Rm(t)-Rf(t)) + c' * Rf(t) + e(t)

Heteroscedasticity is not ruled out by CAPM theory, and in fact, heteroscedasticity is often found to be significant when applying such equations. Therefore, any test of a=0 and c=1 should recognize the possibility of heteroscedasticity.

A SAS dataset with the monthly Fama-French factors is used in this exercise and the data covers the 1965-2004 period (40 years, or 480 monthly observations). The Fama-French factors are periodically updated on Ken French's web site at Darthmouth (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and this set is based on a download made in December 2005. See the WRDS web query for the Fama-French Factors (https://wrds1.wharton.upenn.edu/ds/famafrench/index.shtml) for updated data.

The PROC REG code for estimating SMB and HML equations that assume c=1 (by subtracting RF from SMB and HML) is shown below.

```
Title1 'CAPM type regression equation applied to Fama-French factors';
Title2 'PROC REG OLS fit';
data ffm1;
   set ff.factors_monthly;
   where year(date) between 1985 and 2004;
   *Subtract risk free rate from FF factors, SMB and HML;
   smbrf= smb-rf;
   hmlrf= hml-rf;
run;

proc reg data=ffm1;
   model smbrf = mktrf / spec;
   model hmlrf = mktrf / spec;
   *Include White specification test (SPEC) for heteroscedasticity;
run ;
```

The output from this program for 'CAPM equations applied to Fama-French factors-- PROC REG OLS fit' shows that the intercept for SMB factor is significant, while the intercept for the HML factor is not significant, at least according to conventional t-statistics (-2.16 and 0.63 respectively). Both CAPM betas (MKTRF coefficients) are significantly different from both zero and far below one. However, the White specification test for homoscedasticity versus heteroscedasticity rejects the null hypothesis of homoscedasticity for both equations using a 0.10 critical value ((with a p-values of .007 and 0.060). In other words, using the standard OLS t-statistics is problematic and it would be important to control for heteroscedasticity before drawing any conclusions from the estimated equations.

View Output from PROC REG (SASOutput0.txt)

PROC REG and Heteroscedasticity-Consistent (HC) Covariance Matrices

The ACOV option in PROC REG will add the HC covariance matrix to the output. Code for applying this option is shown below. Although there are two explanatory variables (SMB and HML), a single model statement can be used to estimate two distinct equations because the right-hand side variables are the same for both equations.

```
Title1 'CAPM type regression on Fama-French factors';
Title2 'PROC REG OLS fit with White Correction (ACOV) to Covariance Matrix';
proc reg data=ffm1;
   model smb hml = mktrf rf / spec acov;
   test intercept;
   test intercept=0, rf=1;
run ;
```

In this case, the risk-free rate (RF) is included as a freely estimated right-hand side variable. Note also that the White specification test (SPEC) for heteroscedasticity is selected and that two tests for the coefficients of the equations are specified. The first coefficient test is a simple one that is based on a null hypothesis that the intercept is zero. In some sense, this test is a repetition of the t-statistic for the intercept in each equation, but the output shows that PROC REG treats such a test differently when ACOV is used. The second coefficient test is a joint hypothesis that the intercept is zero and the RF coefficient is one.

The output is shown in the link below. Homoscedasticity is easily rejected for the SMB equation but is not so easily rejected for the HML equation. The effects of the heteroscedasticity correction on the intercept tests are not great, but it is interesting to find that p-value for accepting the joint CAPM restriction (intercept=0, rf=1) falls for both equations (from .0018 to <.0001 for SMB, and from .1691 to .1335 for HML).

It is important to know, however that the HC matrix corrected tests, labeled as 'Results using ACOV estimates' in the PROC REG output are asymptotic-based and thus do not use a sample size or 'degrees of freedom' adjustment that would decrease the test statistic value (slightly in most cases, depending on the number of model parameters). Either PROC SURVEYREG or PROC MODEL must be selected to produce a test that uses of these types of corrections.

View Output from PROC REG using ACOV option (SASHCOutput1.txt)

PROC SURVEYREG and the HC Covariance Matrix

The HC covariance matrix is used in PROC SURVEYREG, such that no additional option must be set. The coefficients are exactly the same as PROC REG produces, but the output shows standard errors that are based on the HC covariance matrix.

```
Title1 'CAPM type regression on Fama-French factors';
Title2 'PROC SURVEYREG with OLS fit and HC Covariance Matrix';
* SMB run-- use excess return (less RF) dependent variable;
proc surveyreg data=ffm1;
   model smbrf = mktrf rf;
   contrast 'smb-test' intercept 1, rf 1;
```

```
run;

* HML run-- use excess return (less RF) dependent variable;
proc surveyreg data=ffm1;
   model hmlrf = mktrf rf;
   contrast 'hml-test' intercept 1, rf 1;
run;
```

PROC SURVEYREG may look to be a generalization of PROC REG, but the syntax is actually based on PROC GLM and two separate procedures must run for each equation. Tests of the coefficients require a CONTRAST statement, which is not as flexible as a TEST statement in PROC REG. For example, PROC SURVEYREG can not simply test the hypothesis that the coefficient of RF is different from 1 in the equations: SMB = a + b*MKTRF + c*RF and HML = a + b*MKTRF + c*RF. Instead, RF must be subtracted from SMB and HML respectively so that an equivalent test can be computed based on the hypothesis that the coefficient of RF in the transformed equations is different from zero. PROC SURVEYREG can handle panel data, as it allows for both CLASS and CLUSTER statements, but it is not as flexible as either PROC GLM or PROC MIXED in this respect.

Output from PROC SURVEYREG (SASHCOutput2.txt)

PROC MODEL and HC Covariance Matrices

PROC MODEL is one of the most flexible and general statistical procedures in SAS (actually part of the SAS/ETS package). To compute a HC covariance matrix use the HCCME option to FIT each equation as below.

```
Title1 'CAPM type regression on Fama-French factors';
Title2 'PROC MODEL OLS fit with HCCME=1 Covariance Matrix';
proc model data=ffm1;
   smb = a1 + b1*mktrf +c1*rf;
   hml = a2 + b2*mktrf +c2*rf;
   fit smb hml / ols hccme=1;
   test a1=0, c1=1;
   test a2=0, c2=1;
run;
```

The HCCME option specifies the type of heteroscedasticity-consistent covariance matrix estimator to use. HCMME=1 makes a degrees of freedom correction, while HCCME=0 will produce a HC covarience matrix that is analagous to using ACOV in PROC REG (i.e. no such correction). Other options are HCCME=2 and HCCME=3 that use corrections suggested by MacKinnon and White (1985) that reduce the effects of outliers and supposedly have better finite sample properties.

Prior to SAS V9 and the addition of the HCCME option, the only way to produce a HC covariance matrix with PROC MODEL was to use GMM estimation as below.

```
Title2 'GMM based HC Covariance Matrix';
proc model data=ffm1;
   smb = a1 + b1*mktrf +c1*rf;
   hml = a2 + b2*mktrf +c2*rf;
   fit smb hml / gmm kernel=(bart, 1 , 0 );
   test a1=0, c1=1;
   test a2=0, c2=1;
   test a1=0, a2=0;
run;
```

GMM estimation typically makes use of instruments, and in this case, the lack of an INSTRUMENT statement means that the explanatory variables serve as their own instruments. Thus the coefficients are the same as the OLS estimates.

The KERNEL setting of (BART,1,0) creates the smallest possible 'window' that resticts the covariance matrix to a case where only contemporaneous correlation among the residuals is used for computing the HC covariance matrix. To make a Newey-West HAC correction, set KERNEL=BART to select the lag length automatically (depending on the sample size) or use KERNEL=(BART,X,0), where X= the lag length + 1 . Other KERNEL settings for PARZEN and QS windows are described in the SAS Documentation for PROC MODEL.

There is also a VARDEF option to specify the denominator to be used in computing the covariance matrix. VARDEF=N uses the number of nonmissing observations and is anlaogous to HCCME=0 and PROC REG's ACOV option. The default is VARDEF=DF, which uses a degrees of freedom correction (the number of nonmissing observations minus the model degrees of freedom).

Note that the PROC MODEL code using GMM above is set up to estimate a system of equations and make use of cross-equation correlation in the error terms. In this application, however, each equation has the same explanatory variables and there is no with-in equation gain in effeciency from doing so. In other words, the same t-statistics and test result for the restriction a1=0 and c1=1 would be produced by running each equation in a separate PROC MODEL block or using FIT with OLS and HCCME=1. Using a single PROC MODEL block and GMM estimation handles cross equation tests in a manner that accounts for the cross equation correlation, however. The test above of a1=0 and a2=0 is an example of such a test.

Output from PROC MODEL (SASHCOutput3.txt)

References

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions" Proceeding of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, 221-233.

MacKinnon, J.G. and H. White (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators With Improved Finite Sample Properties", Journal of Econometrics, 29, 53-57.
MacKinnon, J.G. and H. White (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators With Improved Finite Sample Properties", Journal of Econometrics, 29, 53-57.
SAS Institute Inc. (2004), SAS/ETS User's Guide, Version 9, Cary , NC : SAS Institute Inc.

White, H. (1980), “A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity," Econometrica, 48, 817-838. W