Model Mis-specification

Carlo Favero

Model Mis-specification

- Each specification can be interpreted of the result of a reduction process, what happens if the reduction process that has generated $E(y \mid \mathbf{X})$ omits some relevant information?
- We shall consider three general cases of mis-specification.
  - Mis-specification related to the choice of variables included in the regressions,
  - mis-specification related to ignoring the existence on constraints on the estimated parameters
  - and misspecification related to wrong assumprions on the properties of the error terms.

Choice of variables

- under-parameterization (the estimated model omits variables included in the DGP)
- over-parameterization (the estimated model includes more variables than the DGP).

Under-parameterization

Given the DGP:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{1}$$

for which hypotheses (**??**) – (**??**) hold, the following model is estimated:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\nu}. \tag{2}$$

The OLS estimates are given by the following expression:

$$\widehat{\boldsymbol{\beta}}_1^{up} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}, \tag{3}$$

Under-parameterization

while the OLS estimates which are obtained by estimation of the DGP, are:

$$\widehat{\beta}_1 = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}. \tag{4}$$

The estimates in (4) are best linear unbiased estimators (BLUE) by construction, while the estimates in (3) are biased unless $\mathbf{X}_1$and $\mathbf{X}_2$ are uncorrelated. To show this, consider:

$$\begin{aligned} \widehat{\beta}_1 &= \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\left(\mathbf{X}_1'\mathbf{y} - \mathbf{X}_1'\mathbf{X}_2\widehat{\beta}_2\right) && (5)\\ &= \widehat{\beta}_1^{up} + \widehat{\mathbf{D}}\widehat{\beta}_2, && (6) \end{aligned}$$

where $\widehat{\mathbf{D}}$ is the vector of coefficients in the regression of $\mathbf{X}_2$ on $\mathbf{X}_1$ and $\widehat{\beta}_2$ is the OLS estimator obtained by fitting the DGP.

Clicker 10

Insert Clicker 10 here

Under-parameterization

Note that if

$$E(\mathbf{y} \mid \mathbf{X}_1, \mathbf{X}_2) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,$$
$$E(\mathbf{X}_1 \mid \mathbf{X}_2) = \mathbf{X}_1\mathbf{D},$$

then,

$$E(\mathbf{y} \mid \mathbf{X}_1) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{D}\boldsymbol{\beta}_2 = \mathbf{X}_1\boldsymbol{\alpha}.$$

Therefore the OLS estimator in the under-parameterized model is a biased estimator of $\boldsymbol{\beta}_1$, but an unbiased estimator of $\boldsymbol{\alpha}$.

- Then, if the objective of the model is forecasting and $\mathbf{X}_1$ is more easily observed than $\mathbf{X}_2$, the under-parameterized model can be safely used.
- On the other hand, if the objective of the model is to test specific predictions on parameters, the use of the under-parameterized model delivers biased results.

Over-parameterization

Given the DGP,

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \tag{7}$$

for which hypotheses (**??**) − (**??**) hold, the following model is estimated:

$$\mathbf{y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{v}. \tag{8}$$

The OLS estimator of the over-parameterized model is

$$\widehat{\boldsymbol{\beta}}_1^{op} = \left(\mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1 \right)^{-1} \mathbf{X}_1' \mathbf{M}_2 \mathbf{y}, \tag{9}$$

Over-parameterization

while, by estimating the DGP, we obtain:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}. \tag{10}$$

By substituting $\mathbf{y}$ from the DGP, one finds that both estimators are unbiased and the difference is now made by the variance. In fact we have:

$$\begin{aligned} var\left(\widehat{\boldsymbol{\beta}}_1^{op} \mid \mathbf{X}_1, \mathbf{X}_2\right) &= \sigma^2\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}, \quad (11) \\ var\left(\widehat{\boldsymbol{\beta}}_1 \mid \mathbf{X}_1, \mathbf{X}_2\right) &= \sigma^2\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}. \quad (12) \end{aligned}$$

Over-parameterization

Remember that if two matrices $\mathbf{A}$ and $\mathbf{B}$ are positive definite and $\mathbf{A} - \mathbf{B}$ is positive semidefinite, then also the matrix $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive semidefinite. We have to show that $\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ is a positive semidefinite matrix. Such a result is almost immediate:

$$\begin{aligned}\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1 &= \mathbf{X}_1'\left(\mathbf{I} - \mathbf{M}_2\right)\mathbf{X}_1 \\ &= \mathbf{X}_1'\mathbf{Q}_2\mathbf{X}_1 = \mathbf{X}_1'\mathbf{Q}_2\mathbf{Q}_2\mathbf{X}_1.\end{aligned}$$

We conclude that over-parameterization impacts on the efficiency of estimators and the power of the tests of hypotheses.

Estimation under linear constraints

The estimated model is the linear model analysed up to now:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

while the DGP is instead:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \textit{subject to } \mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0},$$

where the constraints are expressed using the so called implicit form.

Estimation under linear constraints

A useful alternative way of expressing constraints, known as the 'explicit form' has been expressed by Sargan (1988):

$$\boldsymbol{\beta} = \mathbf{S}\boldsymbol{\theta} + \mathbf{s},$$

where $\mathbf{S}$ is a $(k \times (k-r))$ matrix of rank $k-r$ and $\mathbf{s}$ is a $k \times 1$ vector. To show how constraints are specified in the two alternatives let us consider the case of $\beta_1 = -\beta_2$ on the following specification:

$$\ln y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i. \tag{13}$$

Estimation under linear constraints

Using $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$:

$$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = (0),$$

while using $\boldsymbol{\beta} = \mathbf{S}\boldsymbol{\theta} + \mathbf{s}$:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

In practice the constraints in the explicit form are written by considering $\boldsymbol{\theta}$ as the vector of free parameters.

Estimation under linear constraints

Note that there is no unique way of expressing constraints in the explicit form, in our case the same constraint can be imposed as:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

As the two alternatives are indifferent, $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ and $\mathbf{RS}\boldsymbol{\theta} + \mathbf{Rs} - \mathbf{r} = \mathbf{0}$ are equivalent, which implies:

1. $\mathbf{RS} = \mathbf{0}$;
2. $\mathbf{Rs} - \mathbf{r} = \mathbf{0}$.

The restricted least squares (RLS) estimator

To construct RLS, substitute the constraint in the original model to obtain:

$$\mathbf{y} - \mathbf{Xs} = \mathbf{XS}\boldsymbol{\theta} + \boldsymbol{\epsilon}. \tag{14}$$

Equation (14) is equivalent to:

$$\mathbf{y}^{*} = \mathbf{X}^{*}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{15}$$

where $\mathbf{y}^{*} = \mathbf{y} - \mathbf{Xs}$, $\mathbf{X}^{*} = \mathbf{XS}$.
Note that the transformed model features the same residuals with the original model; therefore, if standard hypotheses hold for the original model, they also hold for the transformed.

Clicker 11

Insert Clicker 11 here

The restricted least squares (RLS) estimator

We apply OLS to the transformed model to obtain:

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &= \left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*} \\ &= \left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{X}'\left(\mathbf{y}-\mathbf{X}\mathbf{s}\right). \end{aligned} \tag{16}$$

From (16) the RLS estimation is easily obtained by applying the transformation $\widehat{\boldsymbol{\beta}}^{rls} = \mathbf{S}\widehat{\boldsymbol{\theta}} + \mathbf{s}$. Similarly, the variance of the RLS estimator is easily obtained as:

$$\begin{aligned} var\left(\widehat{\boldsymbol{\theta}} \mid \mathbf{X}\right) &= \sigma^2\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1} = \sigma^2\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}, \\ var\left(\widehat{\boldsymbol{\beta}}^{rls} \mid \mathbf{X}\right) &= var\left(\mathbf{S}\widehat{\boldsymbol{\theta}} + \mathbf{s} \mid \mathbf{X}\right) \\ &= \mathbf{S}\, var\left(\widehat{\boldsymbol{\theta}} \mid \mathbf{X}\right)\mathbf{S}' \\ &= \sigma^2\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'. \end{aligned}$$

The restricted least squares (RLS) estimator

We can now discuss the properties of OLS and RLS in the case of a DGP with constraints.

Unbiasedness
Under the assumed DGP, both estimators are unbiased, since such properties depend on the validity of hypotheses (**??**) − (**??**), which is not affected by the imposition of constraints on parameters.

Efficiency
Obviously, if we interpret RLS as the OLS estimator on the transformed model (16) we immediately derive the results that the RLS is the most efficient estimator, as the hypotheses for the validity of the Gauss Markov theorem are satisfied when OLS is applied to (16). Note that by posing $\mathbf{L} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ in the context of the transformed model, we do not generally obtain OLS but an estimator whose conditional variance with respect to $\mathbf{X}$, coincides with the conditional variance of the OLS estimator.

The restricted least squares (RLS) estimator

We support this intuition with a formal argument

$$var\left(\widehat{\beta} \mid \mathbf{X}\right) - var\left(\widehat{\beta}^{rls} \mid \mathbf{X}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \sigma^2 \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'.$$

Define $\mathbf{A}$ as:

$$\mathbf{A} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'.$$

Given that

$$\begin{aligned}
\mathbf{A}\mathbf{X}'\mathbf{X}\mathbf{A} &= \left(\left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\right)\mathbf{X}'\mathbf{X}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\right. \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1} - 2\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}' + \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1} \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}' \\
&= \mathbf{A},
\end{aligned}$$

$\mathbf{A}$ is positive semidefinite, being the product of a matrix and its transpose.

Heteroscedasticity, Autocorrelation, and the GLS estimator

Let us reconsider the single equation model and generalize it to the case in which the hypotheses of diagonality and constancy of the conditional variances-covariance matrix of the residuals do not hold:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim n.d.\left(\mathbf{0}, \sigma^2 \Omega\right), \end{aligned} \tag{17}$$

where Ω is a $(T \times T)$ symmetric and positive definite matrix. When the OLS method is applied to model (17), it delivers estimators which are consistent but not efficient; moreover, the traditional formula for the variance-covariance matrix of the OLS estimators, $\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$, is wrong and leads to an incorrect inference.

Heteroscedasticity, Autocorrelation, and the GLS estimator

Using the standard algebra, it can be shown that the correct formula for the variance-covariance matrix of the OLS estimator is:

$$\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Omega\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

To find a general solution to this problem, remember that the inverse of a symmetric definite positive matrix is also symmetric and definite positive and that for a given matrix Φ, symmetric and definite positive, there always exists a $(T \times T)$ non-singular matrix $\mathbf{K}$**,** such that $\mathbf{K}'\mathbf{K} = \Omega^{-1}$ and $\mathbf{K}\Omega\mathbf{K}' = \mathbf{I}_T$.

Heteroscedasticity, Autocorrelation, and the GLS estimator

Consider the regression model obtained by pre-multiplying both the right-hand and the left-hand sides of (17) by $\mathbf{K}$:

$$\begin{aligned} \mathbf{Ky} &= \mathbf{KX}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\epsilon}, \\ \mathbf{K}\boldsymbol{\epsilon} &\sim n.d.\left(\mathbf{0}, \sigma^2 \mathbf{I}_T\right). \end{aligned} \tag{18}$$

The OLS estimator of the parameters of the transformed model (18) satisfies all the conditions for the applications of the Gauss−Markov theorem; therefore, the estimator

$$\begin{aligned} \widehat{\beta}_{GLS} &= \left(\mathbf{X}'\mathbf{K}'\mathbf{KX}\right)^{-1}\mathbf{X}'\mathbf{K}'\mathbf{Ky} \\ &= \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}, \end{aligned}$$

known as the generalised least squares (GLS) estimator, is BLUE. The variance of the GLS estimator, conditional upon $\mathbf{X}$, becomes

$$Var\left(\widehat{\beta}_{GLS} \mid \mathbf{X}\right) = \Sigma_{\beta} = \sigma^2 \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}.$$

Heteroscedasticity, Autocorrelation, and the GLS estimator

Consider, for example, the variances of the OLS and the GLS estimators. Using the fact that if **A** and **B** are positive definite and $\mathbf{A}-\mathbf{B}$ is positive semidefinite, then $\mathbf{B}^{-1}-\mathbf{A}^{-1}$ is also positive semidefinite, we have:

$$
\begin{aligned}
& \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right) - \left(\mathbf{X}'\mathbf{X}\right)\left(\mathbf{X}'\Omega\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right) \\
= & \ \mathbf{X}'\mathbf{K}'\mathbf{K}\mathbf{X} - \left(\mathbf{X}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{K}^{-\mathbf{1}}\left(\mathbf{K}'\right)^{-\mathbf{1}}\mathbf{X}\right)^{-\mathbf{1}}\left(\mathbf{X}'\mathbf{X}\right) \\
= & \ \mathbf{X}'\mathbf{K}'\left(\mathbf{I} - \left(\mathbf{K}'\right)^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{K}^{-\mathbf{1}}\left(\mathbf{K}'\right)^{-\mathbf{1}}\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{K}^{-1}\right)\mathbf{K}\mathbf{X} \\
= & \ \mathbf{X}'\mathbf{K}'\mathbf{M}_W'\mathbf{M}_W\mathbf{K}\mathbf{X},
\end{aligned}
$$

where

$$
\mathbf{M}_W = \left(\mathbf{I} - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-\mathbf{1}}\mathbf{W}'\right), \qquad (19)
$$

$$
\mathbf{W} = \left(\mathbf{K}'\right)^{-1}\mathbf{X}. \qquad (20)
$$

Heteroscedasticity, Autocorrelation, and the GLS estimator

- The applicability of the GLS estimator requires an empirical specification for the matrix $\mathbf{K}$.
- We consider here three specific applications where the appropriate choice of such a matrix leads to the solution of the problems in the OLS estimator generated, respectively,
  - by the presence of first-order serial correlation in the residuals,
  - by the presence of heteroscedasticity in the residuals
  - by the presence of both of them.

Correction for Serial Correlation (Cochrane-Orcutt)

Consider first the case of first-order serial correlation in the residuals. We have the following model:

$$\begin{aligned} y_t &= \mathbf{x}_t' \boldsymbol{\beta} + u_t, \\ u_t &= \rho u_{t-1} + \epsilon_t, \\ \epsilon_t &\sim n.i.d. \left(0, \sigma_\epsilon^2\right), \end{aligned}$$

which, using our general notation, can be re-written as:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, && (21) \\ \boldsymbol{\epsilon} &\sim n.d. \left(\mathbf{0}, \sigma^2 \Omega\right), \\ \sigma^2 &= \frac{\sigma_\epsilon^2}{1 - \rho^2}, && (22) \end{aligned}$$

Correction for Serial Correlation (Cochrane-Orcutt)

$$\breve{\ } = \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{T-1} \\ \rho & 1 & \rho & . & . & \rho^{T-2} \\ \rho^2 & . & 1 & . & . & . \\ . & . & . & . & . & . \\ \rho^{T-2} & . & . & \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & . & . & \rho & 1 \end{bmatrix}.$$

In this case, the knowledge of the parameter ρ allows the empirical implementation of the GLS estimator. An intuitive procedure to implement the GLS estimator can then be the following:

1. estimate the vector $\boldsymbol{\beta}$ by OLS and save the vector of residuals $\widehat{u}_t$;
2. regress $\widehat{u}_t$ on $\widehat{u}_{t-1}$ to obtain an estimate $\widehat{\rho}$ of ρ;
3. construct the transformed model and regress $(y_t - \widehat{\rho} y_{t-1})$ on $(\mathbf{x}_t - \widehat{\rho}\mathbf{x}_{t-1})$ to obtain the GLS estimator of the vector of parameters of interest.

Note that the above procedure, known as the Cochrane$-$Orcutt procedure, can be repeated until convergence.

Correction for Heteroscedasticity (White)

In the case of heteroscedasticity, our general model becomes

$$\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \\
\boldsymbol{\epsilon} &\sim n.d.\left(\mathbf{0}, \Omega\right), \\
\Omega &= \begin{bmatrix}
\sigma_1^2 & 0 & 0 & . & . & 0 \\
0 & \sigma_2^2 & 0 & . & . & 0 \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
0 & . & . & 0 & \sigma_{T-1}^2 & 0 \\
0 & 0 & . & . & 0 & \sigma_T^2
\end{bmatrix}.
\end{aligned}$$

In this case, to construct the GLS estimator, we need to model heteroscedasticity choosing appropriately the **K** matrix.

Correction for Heteroscedasticity (White)

White (1980) proposes a specification based on the consideration that in the case of heteroscedasticity the variance-covariance matrix of the OLS estimator takes the form:

$$\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Omega\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1},$$

which can be used for inference, once an estimator for Φ is available. The following unbiased estimator of Φ is proposed:

$$\widehat{\Omega} = \begin{bmatrix} \widehat{u}_1^2 & 0 & 0 & . & . & 0 \\ 0 & \widehat{u}_2^2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & 0 & \widehat{u}_{T-1}^2 & 0 \\ 0 & 0 & . & . & 0 & \widehat{u}_T^2 \end{bmatrix}.$$

Correction for Heteroscedasticity (White)

This choice for $\widehat{\Omega}$ leads to the following degrees of freedom corrected heteroscedasticity consistent parameters' covariance matrix estimator:

$$\Sigma_{\beta}^{W} = \frac{T}{T-k}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\sum_{t=1}^{T}\widehat{u}_t^2\mathbf{x}_t\mathbf{x}_t'\right)\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

This estimator corrects for the OLS for the presence of heteroscedasticity in the residuals without modelling it.

Correction for heteroscedasticity and serial correlation (Newey-West)

The White covariance matrix assumes serially uncorrelated residuals. Newey and West(1987) have proposed a more general covariance estimator that is robust to heteroscedasticity and autocorrelation of the residuals of unknow form. This HAC (heteroscedasticity and autocorrelation consistent) coefficient covariance estimators is given by:

$$\Sigma_{\beta}^{NW} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} T\hat{\Omega} \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

where $\hat{\Omega}$ is a long-run covariance estimators

Correction for heteroscedasticity and serial correlation (Newey-West)

$$\widehat{\Omega} = \widehat{\Gamma}(0) + \sum_{j=1}^{p} \left[1 - \frac{j}{p+1}\right] \left[\widehat{\Gamma}(j) + \widehat{\Gamma}(-j)\right], \tag{23}$$

$$\widehat{\Gamma}(j) = \left(\sum_{t=1}^{T} \widehat{u}_t \widehat{u}_{t-j} \mathbf{X}_t \mathbf{X}'_{t-j}\right) \frac{1}{T}$$

note that in absence of serial correlation $\widehat{\Omega} = \widehat{\Gamma}(0)$ and we are back to the White Estimator. Implementation of the estimator requires a choice of p, which is the maximum lag at whcih correlation is still present. The weighting scheme adopted guarantees a positive definite estimated covariance matrix by multiplying the sequence of the $\widehat{\Gamma}(j)$'s by a sequence of weights that decreases as $|j|$ increases.

Clicker 12

Insert Clicker 12 here

Fama-MacBeth(1973)

- Consider the 25 portfolios and run for each of them the CAPM regression over the sample 1962:1 2014:6. These regressions deliver 25 betas.
- Take now a second-step regression in which the cross-section of the average (over the sample 1962:1-2014:6) monthly returns on the 25 portfolios are projected on the 25 betas:

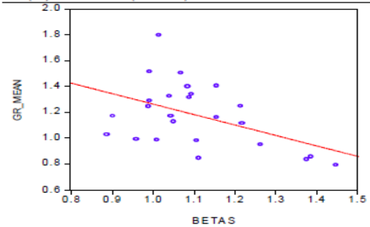
$$r_i = \gamma_0 + \gamma_1 \beta_i + u_i$$

Under the null of the CAPM i) residuals should be randomly distributed around the regression line, ii) $\gamma_0 = E\left(r^f\right), \gamma_1 = E\left(r^m - r^f\right)$

Fama-MacBeth(1973)

TABLE 3.4: CAPM in the cross-section of 25 portfolios

Dependent Variable r_i $(i = 1, ...25)$					
Variable	Coefficient	Std. Error	t-ratio	HAC Std.Err.	Prob.
C	2.07	0.35	5.94	0.451	0.3069
β_i	-0.80	0.31	-2.57	0.364	0.0169
R^2 0.22	S.E. of reg 0.22	S.E. dep.var 0.25	1926:7-2014:6		
$E\left(r^f\right) = 0.40$, $E\left(r^m - r^f\right) = 0.89$					



Fama-MacBeth(1973)

- The cross-sectional regression strongly rejects the CAPM.
- Note however that this regression is affected by an inference problem caused by the correlation of residuals in the cross-section regression.
- Fama and MacBeth (1973) address this problem by estimating month-by-month cross-section regressions of monthly returns on the betas obtained on the full sample.
- The time series means of the monthly slopes and intercepts, along with the standard errors of the means, are then used to run the appropriate tests

Fama-MacBeth(1973)

- The application of the Fama-MacBeth on the sample 1962:1 2014:6 delivers the following results

TABLE 3.5: Statistics on the distribution of coefficients from Fama-MacBeth

	γ_0	γ_1
Mean	2.07	-0.807
St.Dev	9.56	10.06
Obs	630	630
t-stat	5.437	-2.01

An alternative route would be to construct Heteroscedasticity and Autocorrelation Consistent(HAC) estimators.

Clicker 13

Insert Clicker 13 here