Introduction to Econometrics (4th Updated Edition)

by

James H. Stock and Mark W. Watson

Solutions to Odd-Numbered End-of-Chapter Exercises: Chapter 19

(This version September 18, 2018)
19.1. (a) The regression in the matrix form is

\[ Y = X\beta + U \]

with

\[
Y = \begin{pmatrix} 
\text{TestScore}_1 \\
\text{TestScore}_2 \\
\vdots \\
\text{TestScore}_n 
\end{pmatrix}, \quad 
X = \begin{pmatrix} 
1 & \text{Income}_1 & \text{Income}_1^2 \\
1 & \text{Income}_2 & \text{Income}_2^2 \\
\vdots & \vdots & \vdots \\
1 & \text{Income}_n & \text{Income}_n^2 
\end{pmatrix}
\]

\[
U = \begin{pmatrix} 
U_1 \\
U_2 \\
\vdots \\
U_n 
\end{pmatrix}, \quad 
\beta = \begin{pmatrix} 
\hat{\beta}_0 \\
\hat{\beta}_1 \\
\hat{\beta}_2 
\end{pmatrix}
\]

(b) The null hypothesis is

\[ R\beta = r \]

versus \( R\beta \neq r \) with

\[ R = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \text{ and } r = 0. \]

The heteroskedasticity-robust \( F \)-statistic testing the null hypothesis is

\[
F = (R\hat{\beta} - r)' \left[ R\Sigma R' \right]^{-1} (R\hat{\beta} - r) / q
\]

With \( q = 1 \). Under the null hypothesis,

\[ F \xrightarrow{d} F_{q,\infty}. \]

We reject the null hypothesis if the calculated \( F \)-statistic is larger than the critical value of the \( F_{q,\infty} \) distribution at a given significance level.
19.3. (a)

\[
\text{Var}(Q) = E[(Q - \mu_Q)^2] \\
= E[(Q - \mu_Q)(Q - \mu_Q)'] \\
= E[(c'W - c'\mu_w)(c'W - c'\mu_w)'] \\
= c'E[(W - \mu_w)(W - \mu_w)']c \\
= c' \text{var}(W)c = c'\Sigma_w c
\]

where the second equality uses the fact that \( Q \) is a scalar and the third equality uses the fact that \( \mu_Q = c'\mu_w \).

(b) Because the covariance matrix \( \Sigma_w \) is positive definite, we have \( c'\Sigma_w c > 0 \) for every non-zero vector from the definition. Thus, \( \text{var}(Q) > 0 \). Both the vector \( c \) and the matrix \( \Sigma_w \) are finite, so \( \text{var}(Q) = c'\Sigma_w c \) is also finite. Thus, \( 0 < \text{var}(Q) < \infty \).
19.5.  \( P_X = X (X'X)^{-1} X' \), \( M_X = I_n - P_X \).

(a)  \( P_X \) is idempotent because

\[
P_X P_X = X (X'X)^{-1} X X (X'X)^{-1} X' = X (X'X)^{-1} \hat{X} = P_X.
\]

\( M_X \) is idempotent because

\[
M_X P_X = (I_n - P_X) (I_n - P_X) = I_n - P_X - P_X + P_X P_X = I_n - 2P_X + P_X = I_n - P_X = M_X
\]

\( P_X M_X = 0_{n \times n} \) because

\[
P_X M_X = P_X (I_n - P_X) = P_X - P_X P_X = P_X - P_X = 0_{n \times n}
\]

(b) Because \( \hat{\beta} = (X'X)^{-1} X'Y \), we have

\[
\hat{Y} = X \hat{\beta} = X (X'X)^{-1} X'Y = P_X Y
\]

which is Equation (19.27). The residual vector is

\[
\hat{U} = Y - \hat{Y} = Y - P_X Y = (I_n - P_X) Y = M_X Y.
\]

We know that \( M_X X \) is orthogonal to the columns of \( X \):

\[
M_X X = (I_n - P_X) X = X = P_X X = X + X (X'X)^{-1} X' X - X = 0
\]

so the residual vector can be further written as

\[
\hat{U} = M_X Y = M_X (X \beta + U) = M_X X \beta + M_X U = M_X U
\]

which is Equation (19.28).

(c) From the hint, rank \((P_X) = \text{trace}(P_X) = \text{trace}[X (X'X)^{-1} X'] = \text{trace}[(X'X)^{-1} X' X] = \text{trace}(I_{k+1}) = k+1\). The result for \( M_X \) follows from a similar calculation.
19.7. (a) We write the regression model, \( Y_i = \beta_1 X_i + \beta_2 W_i + u_i \), in the matrix form as

\[
Y = X\beta_1 + W\beta_2 + U
\]

with

\[
Y = \begin{pmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_n
\end{pmatrix}, \quad X = \begin{pmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{pmatrix}, \quad W = \begin{pmatrix}
W_1 \\
W_2 \\
\vdots \\
W_n
\end{pmatrix}, \quad U = \begin{pmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{pmatrix},
\]

The OLS estimator is

\[
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{pmatrix} = \left( X'X \quad X'W \right) \left( X'Y \right)
\]

\[
= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \left( \frac{1}{n} X'X \quad \frac{1}{n} X'W \right)^{-1} \left( \frac{1}{n} X'U \right)
\]

\[
= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 \quad \frac{1}{n} \sum_{i=1}^{n} X_i W_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} X_i u_i \right)
\]

By the law of large numbers, \( \frac{1}{n} \sum_{i=1}^{n} X_i^2 \xrightarrow{d} E(X^2) \); \( \frac{1}{n} \sum_{i=1}^{n} W_i^2 \xrightarrow{d} E(W^2) \);

\( \frac{1}{n} \sum_{i=1}^{n} X_i W_i \xrightarrow{d} E(XW) = 0 \) (because \( X \) and \( W \) are independent with means of zero);

\( \frac{1}{n} \sum_{i=1}^{n} X_i u_i \xrightarrow{d} E(Xu) = 0 \) (because \( X \) and \( u \) are independent with means of zero);

\( \frac{1}{n} \sum_{i=1}^{n} X_i u_i \xrightarrow{d} E(Xu) = 0 \)

Thus

\[
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{pmatrix} \xrightarrow{d} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \left( \frac{1}{n} X'X \quad 0 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} X_i u_i \right)
\]

\[
= \begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix} + \frac{E(Wu)}{E(W^2)}.
\]
(b) From the answer to (a) \( \hat{\beta}_2 \xrightarrow{d} \beta_2 + \frac{E(Wu)}{E(W^2)} \neq \beta_2 \) if \( E(Wu) \) is nonzero.

(c) Consider the population linear regression \( u_i \) onto \( W_i \):

\[
u_i = \lambda W_i + a_i
\]

where \( \lambda = E(Wu)/E(W^2) \). In this population regression, by construction, \( E(aW) = 0 \). Using this equation for \( u_i \) rewrite the equation to be estimated as

\[
Y_i = X_i \beta_1 + W_i \beta_2 + u_i \\
= X_i \beta_1 + W_i (\beta_2 + \lambda) + a_i \\
= X_i \beta_1 + W_i \theta + a_i
\]

where \( \theta = \beta_2 + \lambda \). A calculation like that used in part (a) can be used to show that

\[
\left( \frac{\sqrt{n}(\hat{\beta}_1 - \beta_1)}{\sqrt{n}(\hat{\beta}_2 - \theta)} \right) \xrightarrow{d} \left[ \begin{pmatrix} 1 \frac{\sum W_i^2}{n} \frac{\sum X_i W_i}{n} \\ \frac{\sum W_i}{n} \sum X_i W_i \frac{\sum W_i}{n} \sum W_i^2 \frac{\sum W_i}{n} \sum W_i a_i \end{pmatrix} \right]^{-1} \begin{pmatrix} \frac{1}{n} \sum X_i W_i a_i \\ \frac{1}{n} \sum W_i a_i \end{pmatrix}
\]

\[
\xrightarrow{d} \begin{pmatrix} E(X^2) & 0 \\ 0 & E(W^2) \end{pmatrix}^{-1} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}
\]

where \( S_1 \) is distributed \( N(0, \sigma_a^2 E(X_2)) \). Thus by Slutsky’s theorem

\[
\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N \left( 0, \frac{\sigma_a^2}{E(X^2)} \right)
\]

Now consider the regression that omits \( W \), which can be written as:

\[
Y_i = X_i \beta_1 + d_i
\]

where \( d_i = W_i \theta + a_i \). Calculations like those used above imply that

\[
\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N \left( 0, \frac{\sigma_d^2}{E(X^2)} \right)
\]

Since \( \sigma_d^2 = \sigma_a^2 + \theta^2 E(W^2) \), the asymptotic variance of \( \hat{\beta}_1^* \) is never smaller than the asymptotic variance of \( \hat{\beta}_1 \).
19.9. (a)

\[ \hat{\beta} = (X'M_wX)^{-1}X'M_wY \]
\[ = (X'M_wX)^{-1}X'M_w(X\beta + Wy + U) \]
\[ = \beta + (X'M_wX)^{-1}X'M_wU. \]

The last equality has used the orthogonality \( M_W = 0 \). Thus

\[ \hat{\beta} - \beta = (X'M_wX)^{-1}X'M_wU = (n^{-1}X'M_wX)^{-1}(n^{-1}X'M_wU). \]

(b) Using \( M_W = I_n - P_W \) and \( P_W = W(W'W)^{-1}W' \) we can get

\[ n^{-1}X'M_wX = n^{-1}X'(I_n - P_W)X \]
\[ = n^{-1}XX - n^{-1}XP_WX \]
\[ = n^{-1}XX - (n^{-1}X'W)(n^{-1}W'W)^{-1}(n^{-1}W'X). \]

First consider \( n^{-1}X'X = \frac{1}{n} \sum_{i=1}^{n} X_iX_i' \). The \((j, l)\) element of this matrix is

\( \frac{1}{n} \sum_{i=1}^{n} X_{ji}X_{li} \). By Assumption (ii), \( X_i \) is i.i.d., so \( X_{ji}X_{li} \) is i.i.d. By Assumption (iii) each element of \( X_i \) has four moments, so by the Cauchy-Schwarz inequality \( X_{ji}X_{li} \) has two moments:

\[ E(X_{ji}^2X_{li}^2) \leq \sqrt{E(X_{ji}^4) \cdot E(X_{li}^4)} < \infty. \]

Because \( X_{ji}X_{li} \) is i.i.d. with two moments, \( \frac{1}{n} \sum_{i=1}^{n} X_{ji}X_{li} \) obeys the law of large numbers, so

\[ \frac{1}{n} \sum_{i=1}^{n} X_{ji}X_{li} \xrightarrow{p} E(X_{ji}X_{ji}). \]

This is true for all the elements of \( n^{-1}X'X \), so

\[ n^{-1}X'X = \frac{1}{n} \sum_{i=1}^{n} X_iX_i' \xrightarrow{p} E(X_iX_i') = \Sigma_{XX}. \]

Applying the same reasoning and using Assumption (ii) that \( (X_i, W_i, Y_i) \) are i.i.d. and Assumption (iii) that \( (X_i, W_i, u_i) \) have four moments, we have
\[ n^{-1} WW = \frac{1}{n} \sum_{i=1}^{n} W_i W_i' \Rightarrow E(W_i W_i') = \Sigma_{ww}, \]
\[ n^{-1} X W = \frac{1}{n} \sum_{i=1}^{n} X_i W_i' \Rightarrow E(X_i W_i') = \Sigma_{xw}, \]
and
\[ n^{-1} W' X = \frac{1}{n} \sum_{i=1}^{n} W_i' X_i' \Rightarrow E(W_i' X_i') = \Sigma_{wx}. \]

From Assumption (iii) we know \( \Sigma_{xx}, \Sigma_{ww}, \Sigma_{xw}, \) and \( \Sigma_{wx} \) are all finite non-zero, Slutsky’s theorem implies
\[ n^{-1} X'M_w X = n^{-1} X'X - (n^{-1} X'W)(n^{-1} W'W)^{-1}(n^{-1} W'X) \]
\[ \Rightarrow \Sigma_{xx} - \Sigma_{xw} \Sigma_{ww}^{-1} \Sigma_{wx} \]
which is finite and invertible.

(c) The conditional expectation
\[
E(U|X, W) = \begin{pmatrix}
E(u_1|X, W) \\
E(u_2|X, W) \\
\vdots \\
E(u_n|X, W)
\end{pmatrix} = \begin{pmatrix}
E(u_1|X_1, W_1) \\
E(u_2|X_2, W_2) \\
\vdots \\
E(u_n|X_n, W_n)
\end{pmatrix}
= \begin{pmatrix}
W_1' \delta \\
W_2' \delta \\
\vdots \\
W_n' \delta
\end{pmatrix} = \begin{pmatrix}
W_1' \\
W_2' \\
\vdots \\
W_n'
\end{pmatrix} \delta = W \delta.
\]
The second equality used Assumption (ii) that \((X_i, W_i, Y_i)\) are i.i.d., and the third equality applied the conditional mean independence assumption (i).
(d) In the limit

\[ n^{-1}X'M_wU \xrightarrow{p} E(X'M_wU|X, W) = X'M_wE(U|X, W) = X'M_wW\delta = 0 \]

because \( M_wW = 0 \).

(e) \( n^{-1}X'M_wX \) converges in probability to a finite invertible matrix, and 
\( n^{-1}X'M_wU \) converges in probability to a zero vector. Applying Slutsky’s theorem,

\[ \hat{\beta} - \beta = (n^{-1}X'M_wX)^{-1} (n^{-1}X'M_wU) \xrightarrow{p} 0. \]

This implies

\[ \hat{\beta} \xrightarrow{p} \beta. \]
19.11. (a) Using the hint \( C = [Q_1 \ Q_2] \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} [Q_1^\prime \ Q_2^\prime] \), where \( Q^\prime Q = I \). The result follows with \( A = Q_1 \).

(b) \( W = A^\prime V \sim N(A^\prime 0, A^\prime I_n A) \) and the result follows immediately.

(c) \( V^\prime CV = V^\prime AA^\prime V = (A^\prime V)(A^\prime V) = W^\prime W \) and the result follows from (b).
19.13. (a) This follows from the definition of the Lagrangian.

(b) The first order conditions are


\[
(*) \ X'(Y - X \hat{\beta}) + R' \lambda = 0
\]

and

\[
(**) \ R \hat{\beta} - r = 0
\]

Solving (*) yields

\[
(***) \hat{\beta} = \beta + (XX)^{-1}R' \lambda.
\]

Multiplying by \( R \) and using (**) yields \( r = R \hat{\beta} + R(XX)^{-1}R' \lambda \), so that

\[
\lambda = -[R(XX)^{-1}R']^{-1}(R \hat{\beta} - r).
\]

Substituting this into (***) yields the result.

(c) Using the result in (b), \( Y - X \hat{\beta} = (Y - X \hat{\beta}) - X(XX)^{-1}R' [R(XX)^{-1}R']^{-1}(R \hat{\beta} - r) \), so that

\[
(Y - X \hat{\beta})'(Y - X \hat{\beta}) = (Y - X \hat{\beta})'(Y - X \hat{\beta}) + (R \hat{\beta} - r)' [R(XX)^{-1}R']^{-1}(R \hat{\beta} - r)
\]

\[
+ 2(Y - X \hat{\beta})' X(XX)^{-1}R' [R(XX)^{-1}R']^{-1}(R \hat{\beta} - r).
\]

But \( (Y - X \hat{\beta})' X = 0 \), so the last term vanishes, and the result follows.

(d) The result in (c) shows that \( (R \hat{\beta} - r)' [R(XX)^{-1}R']^{-1}(R \hat{\beta} - r) = SSR_{Restricted} - SSR_{Unrestricted} \). Also \( s_u^2 = SSR_{Unrestricted}/(n - k_{Unrestricted} - 1) \), and the result follows immediately.
19.15. (a) This follows from exercise (19.6).

(b) \( \hat{Y}_i = \hat{X}_i \beta + \hat{u}_i \), so that

\[
\hat{\beta} - \beta = \left( \sum_{i=1}^{n} \hat{X}_i' \hat{X}_i \right)^{-1} \sum_{i=1}^{n} \hat{X}_i' \hat{u}_i
\]

\[
= \left( \sum_{i=1}^{n} \hat{X}_i' \hat{X}_i \right)^{-1} \sum_{i=1}^{n} X_i' M' Mu_i
\]

\[
= \left( \sum_{i=1}^{n} \hat{X}_i' \hat{X}_i \right)^{-1} \sum_{i=1}^{n} X_i' M' u_i
\]

\[
= \left( \sum_{i=1}^{n} \hat{X}_i' \hat{X}_i \right)^{-1} \sum_{i=1}^{n} \hat{X}_i' u_i
\]

(c) \( \hat{Q}_X = \frac{1}{n} \sum_{i=1}^{n} (T^{-1} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)^2) \), where \( (T^{-1} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)^2) \) are i.i.d. with mean \( Q_X \) and finite variance (because \( X_{it} \) has finite fourth moments). The result then follows from the law of large numbers.

(d) This follows the the Central limit theorem.

(e) This follows from Slutsky’s theorem.

(f) \( \eta_i^2 \) are i.i.d., and the result follows from the law of large numbers.

(g) Let \( \hat{\eta}_i = T^{-1/2} \hat{X}_i' \hat{u}_i = \eta_i - T^{-1/2} (\hat{\beta} - \beta) \hat{X}_i' \hat{X}_i \). Then

\[
\hat{\eta}_i^2 = T^{-1/2} \hat{X}_i' \hat{u}_i = \eta_i^2 + T^{-1} (\hat{\beta} - \beta)^2 (\hat{X}_i' \hat{X}_i)^2 - 2T^{-1/2} (\hat{\beta} - \beta) \eta_i \hat{X}_i' \hat{X}_i
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{\eta}_i^2 - \frac{1}{n} \sum_{i=1}^{n} \eta_i^2 = T^{-1} (\hat{\beta} - \beta)^2 \frac{1}{n} \sum_{i=1}^{n} (\hat{X}_i' \hat{X}_i)^2 - 2T^{-1/2} (\hat{\beta} - \beta) \frac{1}{n} \sum_{i=1}^{n} \eta_i \hat{X}_i' \hat{X}_i
\]

Because \( (\hat{\beta} - \beta) \xrightarrow{p} 0 \), the result follows from (a) \( \frac{1}{n} \sum_{i=1}^{n} (\hat{X}_i' \hat{X}_i)^2 \xrightarrow{p} E[(\hat{X}_i' \hat{X}_i)^2] \) and (b) \( \frac{1}{n} \sum_{i=1}^{n} \eta_i \hat{X}_i' \hat{X}_i \xrightarrow{p} E(\eta_i \hat{X}_i' \hat{X}_i) \). Both (a) and (b) follow from the law of large numbers; both (a) and (b) are averages of i.i.d. random variables.

Completing the proof requires verifying that \( (\hat{X}_i' \hat{X}_i)^2 \) has two finite moments.
and $\eta_i \tilde{X}_i', \tilde{X}_i$ has two finite moments. These in turn follow from 8-moment assumptions for $(X_{it}, u_{it})$ and the Cauchy-Schwartz inequality. Alternatively, a “strong” law of large numbers can be used to show the result with finite fourth moments.
19.17 The results follow from the hints and matrix multiplication and addition.