9.7 Dummy Variables

**Definition 9.7.1:** A *dummy (or binary) variable* is a variable which can take on only two values, usually either zero or one.
Example 9.7.1 (Formulation I): Consider a sample of $T$ recent university job market candidates who successfully obtained their first full-time job. For $t = 1, 2, \ldots, T$ define the following variables:

$$y_t = \text{starting earnings of candidate } t,$$
\[ z_{t1} = \begin{cases} 
1, & \text{if the highest degree achieved by candidate } t \text{ is a bachelors} \\
0, & \text{otherwise} 
\end{cases}, \]

\[ z_{t2} = \begin{cases} 
1, & \text{if the highest degree achieved by candidate } t \text{ is a masters} \\
0, & \text{otherwise} 
\end{cases}, \]

\[ z_{t3} = \begin{cases} 
1, & \text{if the highest degree achieved by candidate } t \text{ is a PhD} \\
0, & \text{otherwise} 
\end{cases}. \]
Suppose the $T$ observations are ordered so that the first $T_1$ have $z_{t1} = 1$, the next $T_2$ have $z_{t2} = 1$, and the last $T_3$ have $z_{t3} = 1$, where $T = T_1 + T_2 + T_3$. Under the standard multiple linear normal regression model with fixed regressors, consider

$$y_t = \beta_1 + \beta_2 z_{t2} + \beta_3 z_{t3} + u_t \quad (t = 1, 2, ..., T), \quad (9.7.1a)$$

or

$$y = Z_1 \beta + u, \quad (9.7.1b)$$

where
\[ Z_I = \begin{bmatrix} \mathbf{i}_{T_1}' & \mathbf{i}_{T_2}' & \mathbf{i}_{T_3}' \\ 0_{T_1}' & \mathbf{i}_{T_2}' & 0_{T_3}' \\ 0_{T_1}' & 0_{T_2}' & \mathbf{i}_{T_3}' \end{bmatrix}, \quad (9.7.2) \]

\[ Z_I'Z_I = \begin{bmatrix} T & T_2 & T_3 \\ T_2 & T_2 & 0 \\ T_3 & 0 & T_3 \end{bmatrix}, \quad (9.7.3) \]

\[ (Z_I'Z_I)^{-1} = \begin{bmatrix} \frac{1}{T_1} & -\frac{1}{T_1} & -\frac{1}{T_1} \\ -\frac{1}{T_1} & \frac{T_1+T_2}{T_1T_2} & \frac{1}{T_1} \\ -\frac{1}{T_1} & \frac{1}{T_1} & \frac{T_1+T_3}{T_1T_3} \end{bmatrix}, \quad (9.7.4) \]
\[ Z_1' y = [T \bar{y}, T_2 \bar{y}_2, T_3 \bar{y}_3]' \tag{9.7.5} \]

The OLS estimator is given by

\[ b = [\bar{y}_1, \bar{y}_2 - \bar{y}_1, \bar{y}_3 - \bar{y}_1]' \tag{9.7.6} \]

where \( \bar{y}_1, \bar{y}_2, \) and \( \bar{y}_3 \) are the mean earnings corresponding to the first \( T_1 \), the next \( T_2 \), and the last \( T_3 \) observations.
To aid in the interpretation of parameters, consider:

\[ E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \quad (9.7.7) \]

\[ E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_1 + \beta_2, \quad (9.7.8) \]

\[ E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_1 + \beta_3. \quad (9.7.9) \]

- \( \beta_1 \) is the expected earnings of a candidate with only a bachelors degree.
- \( \beta_2 \) is the expected masters/bachelors earnings differential.
- \( \beta_3 \) is expected PhD/bachelors earnings differential.
\[ E(y_t \mid z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \] (9.7.7)

\[ E(y_t \mid z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_1 + \beta_2, \] (9.7.8)

\[ E(y_t \mid z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_1 + \beta_3. \] (9.7.9)

- Under

\[ H_1: \beta_2 = 0, \] (9.7.10)

holders of a masters as their highest degree earn the same starting earnings as those with a bachelors as their highest degree.
\[ E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \]  \hspace{1cm} (9.7.7)

\[ E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_1 + \beta_2, \]  \hspace{1cm} (9.7.8)

\[ E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_1 + \beta_3. \]  \hspace{1cm} (9.7.9)

- Under

\[ H_2: \beta_3 = 0, \]  \hspace{1cm} (9.7.11)

holders of a PhD as their highest degree earn the same starting earnings as those with a bachelors degree as their highest degree.
\[
E(y_t|z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1,
\] (9.7.7)

\[
E(y_t|z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_1 + \beta_2,
\] (9.7.8)

\[
E(y_t|z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_1 + \beta_3.
\] (9.7.9)

- Under

\[H_3: \beta_2 = \beta_3,\] (9.7.12)

holders of a masters as their highest degree earn the same starting earnings as those with a PhD as their highest degree.
\[ E(y_t \mid z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \quad (9.7.7) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_1 + \beta_2, \quad (9.7.8) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_1 + \beta_3. \quad (9.7.9) \]

- Under

\[ H_4: \beta_2 = \beta_3 = 0, \quad (9.7.13) \]

all three educational groups have the same expected starting earnings.

- Test of joint null hypothesis (9.7.13) is an F-test with 2 degrees of freedom in the numerator.

- Against two-sided alternatives, tests of hypotheses (9.7.10) - (9.7.12) are t-tests.
Note:

- One-sided alternatives $\beta_2 > 0$, $\beta_3 > 0$, $\beta_3 > \beta_2$, seem more appropriate in the present context.
  
  - In the case of (9.7.10) - (9.7.12), these one sided alternatives require standard *one-sided t-tests*.

  - For (9.7.13) a joint one-sided alternative is difficult to handle.
• The case of G mutually exclusive groups is straightforward. It requires G - 1 dummies in addition to the unitary regressor corresponding to the intercept.

• The coefficients of the G - 1 dummies measure deviations in the conditional means of \( y_t \) for the respective groups from the conditional mean for the omitted group.

• If all G dummies are used together with an intercept, then the regressor matrix will not have full column rank.
Example 9.7.2 (Formulation II): Consider

\[ y_t = \beta_1 z_{t1} + \beta_2 z_{t2} + \beta_3 z_{t3} + u_t \quad (t = 1, 2, \ldots, T) \]  \hfill (9.7.14)

or

\[ y = Z_{II} \beta + u, \]  \hfill (9.7.15)

where

\[
Z_{II} = \begin{bmatrix}
1_{T_1}' & 0_{T_2}' & 0_{T_3}' \\
0_{T_1}' & 1_{T_2}' & 0_{T_3}' \\
0_{T_1}' & 0_{T_2}' & 1_{T_3}'
\end{bmatrix}',
\]  \hfill (9.7.16)
\[ Z_{\text{II}'} Z_{\text{II}} = \begin{bmatrix} T_1 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_3 \end{bmatrix} \]  

(9.7.17)

\[
(Z_{\text{II}'} Z_{\text{II}})^{-1} = \begin{bmatrix} \frac{1}{T_1} & 0 & 0 \\ 0 & \frac{1}{T_2} & 0 \\ 0 & 0 & \frac{1}{T_3} \end{bmatrix},
\]

(9.7.18)

\[ Z_{\text{II}'} y = [T_1 \bar{y}_1, T_2 \bar{y}_2, T_3 \bar{y}_3]' .\]  

(9.7.19)
The OLS estimator is given by

\[ b = [\bar{y}_1, \bar{y}_2, \bar{y}_3]', \quad (9.7.20) \]

In Parameterization II levels are estimated directly:

\[ E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \quad (9.7.21) \]
\[ E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_2, \quad (9.7.22) \]
\[ E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_3. \quad (9.7.23) \]

• Unlike Formulation I, null hypotheses $H_1: \beta_2 = 0$ and $H_2: \beta_3 = 0$ are of little concern in Formulation II since zero starting earnings for any of the groups do not constitute interesting assertions.
\[
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1, \quad (9.7.21)
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_2, \quad (9.7.22)
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_3. \quad (9.7.23)
\]

- Null hypothesis \( H_3: \beta_2 = \beta_3 \) remains of interest and has the same interpretation as in Formulation I.

- Null hypothesis \( H_4: \beta_2 = \beta_3 = 0 \) in Formulation I (all three educational groups have the same mean starting earnings, becomes in Formulation II:

\[
H_5: \beta_1 = \beta_2 = \beta_3. \quad (9.7.24)
\]
Example 9.7.3 (Formulation III): Consider

\[ y_t = \beta_1 z_{t1} + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4 + u_t, \quad (t = 1, 2, \ldots, T), \]  

(9.7.25)

subject to

\[ \beta_1 + \beta_2 + \beta_3 = 0. \]  

(9.7.26)

Using (9.7.26) to substitute-out \( \beta_3 \) in (9.7.25) yields
\[ y_t = \beta_1 z_{t1} + \beta_2 z_{t2} + (-\beta_1 - \beta_2)z_{t3} + \beta_4 + u_t \]

\[ = \beta_1 (z_{t1} - z_{t3}) + \beta_2 (z_{t2} - z_{t3}) + \beta_4 + u_t, \quad (9.7.27) \]

where \( \beta_4 \) is a “pure” intercept.
To aid in the interpretation of the parameters in this model, consider

\[
E(y_t \mid z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1 + \beta_4, \quad (9.7.28)
\]

\[
E(y_t \mid z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_2 + \beta_4, \quad (9.7.29)
\]

\[
E(y_t \mid z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_3 + \beta_4. \quad (9.7.30)
\]

Also consider the marginal mean of \( y_t \) for a candidate who is equally likely to have each of the three possible terminal degrees. Then averaging across (9.7.28) - (9.7.30) and noting (9.7.26) yields

\[
E(y_t) = \beta_4. \quad (9.7.31)
\]
\[
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0) = \beta_1 + \beta_4, \quad (9.7.28)
\]
\[
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0) = \beta_2 + \beta_4, \quad (9.7.29)
\]
\[
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1) = \beta_3 + \beta_4. \quad (9.7.30)
\]

- Comparison of (9.7.28)-(9.7.30) with (9.7.31) implies that \( \beta_k \) corresponds to deviation of the expected starting earnings for education group \( k \) (\( k = 1, 2, 3 \)) from that of the “grand mean” in (9.7.31).

- Null hypotheses (9.7.10) - (9.7.12) correspond to deviations from the grand mean rather than the mean of the omitted group as in I.
• Formulations I, II, and III are equivalent one-way analysis of variance models involving the single “factor” education. By adding a second factor (e.g. gender) the model becomes a two-way ANOVA model.
Example 9.7.4: Define $z_t = [z_{t1}, z_{t2}, z_{t3}, z_{t4}]'$, where

$$z_{t4} = \begin{cases} 
1, & \text{if candidate } t \text{ is female} \\
0, & \text{otherwise} 
\end{cases}.$$  \hspace{1cm} (9.7.32)

Consider modeling conditional mean $E(y_t | z_t)$ by adding $z_{t4}$ to Formulation I:

$$y_t = \beta_1 + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4 z_{t4} + u_t, \quad (t = 1, 2, \ldots, T). $$ \hspace{1cm} (9.7.33)

The interpretation of parameters can be obtained by considering
\begin{align*}
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 0) &= \beta_1, \quad (9.7.34) \\
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 0) &= \beta_1 + \beta_2, \quad (9.7.35) \\
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 0) &= \beta_1 + \beta_3, \quad (9.7.36) \\
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 1) &= \beta_1 + \beta_4, \quad (9.7.37) \\
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 1) &= \beta_1 + \beta_2 + \beta_4, \quad (9.7.38) \\
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 1) &= \beta_1 + \beta_3 + \beta_4. \quad (9.7.39)
\end{align*}

- The interpretation of $\beta_1$, $\beta_2$, and $\beta_3$ is the same as Example 9.7.1 for males.
• $\beta_4$ is the additive female differential which is constant across the three educational groups.

• An obvious null hypothesis of interest is

$$H_6: \beta_4 = 0.$$  \hspace{1cm} (9.7.40)
Example 9.7.5: Consider the following \textit{two-way ANOVA model with interaction terms}:

\[ y_t = \beta_1 + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4 z_{t4} + \beta_5 (z_{t2} z_{t4}) + \beta_6 (z_{t3} z_{t4}) + u_t. \] (9.7.41)

To help interpret the parameters, consider:
\[
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 0) = \beta_1, \tag{9.7.42}
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 0) = \beta_1 + \beta_2, \tag{9.7.43}
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 0) = \beta_1 + \beta_3, \tag{9.7.44}
\]

\[
E(y_t | z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 1) = \beta_1 + \beta_4, \tag{9.7.45}
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 1) = \beta_1 + \beta_2 + \beta_4 + \beta_5, \tag{9.7.46}
\]

\[
E(y_t | z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 1) = \beta_1 + \beta_3 + \beta_4 + \beta_6. \tag{9.7.47}
\]
The interpretation of $\beta_1$, $\beta_2$, and $\beta_3$ is the same as Example 9.7.1, except that it now applies only to males.

Null hypotheses (9.7.10) - (9.7.12) are of interest for males.

Unlike Example 9.7.4, in which the female/male earnings differential is constant across the three educational groups, in Example 9.7.5 the female/male earnings differential varies across the three educational groups.

$\beta_4$ is the additive female/male expected earnings differential for those candidates having a bachelors as the highest degree.
\[ \beta_5 \text{ is the additive female/male expected earnings differential on top of the bachelors differential for those candidates having a masters as the highest degree.} \]

\[ \beta_6 \text{ is the additive female/male expected earnings differential on top of the bachelor differential for those candidates having a PhD.} \]

\[ \beta_2 + \beta_5 \text{ and } \beta_3 + \beta_6 \text{ are the MA/BA and PhD/BA expected earnings differentials for females.} \]
• In addition to null hypothesis $H_6: \beta_4 = 0$, it is also interesting to test:

\[
H_7: \beta_5 = 0, \quad (9.7.48)
\]
\[
H_8: \beta_6 = 0. \quad (9.7.49)
\]

• The null hypothesis that there is no female/male expected earnings differential at any educational level corresponds to

\[
H_9: \beta_4 = \beta_5 = \beta_6 = 0. \quad (9.7.50)
\]

• The test that expected starting earnings is the same for all candidates is

\[
H_{10}: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0. \quad (9.7.51)
\]
**Example 9.7.6:** Consider the following alternative formulation:

\[ y_t = \beta_1 z_{t1} + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4 (z_{t1} z_{t4}) + \beta_5 (z_{t2} z_{t4}) + \beta_6 (z_{t3} z_{t4}) + u_t \]

\( (t = 1, 2, ..., T). \quad (9.7.52) \)

The interpretation of parameters can be obtained by considering...
\[ E(y_t \mid z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 0) = \beta_1, \quad (9.7.53) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 0) = \beta_2, \quad (9.7.54) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 0) = \beta_3, \quad (9.7.55) \]
\[ E(y_t \mid z_{t1} = 1, z_{t2} = 0, z_{t3} = 0, z_{t4} = 1) = \beta_1 + \beta_4, \quad (9.7.56) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 1, z_{t3} = 0, z_{t4} = 1) = \beta_2 + \beta_5, \quad (9.7.57) \]
\[ E(y_t \mid z_{t1} = 0, z_{t2} = 0, z_{t3} = 1, z_{t4} = 1) = \beta_3 + \beta_6. \quad (9.7.58) \]
• In (9.7.52) expected earnings levels for males $\beta_k$ ($k = 1, 2, 3$) are estimated directly.

• Null hypotheses $H_1: \beta_2 = 0$ and $H_2: \beta_3 = 0$ are of little interest.

• Null hypothesis $H_3: \beta_2 = \beta_3$ remains of interest and has the same interpretation as in Formulation I, except that it holds only for males.
• Null hypothesis $H_3: \beta_2 = \beta_3 = 0$, all three male educational groups have the same expected starting earnings, becomes

$$H_{11}: \beta_1 = \beta_2 = \beta_3.$$  \hfill (9.7.59)

The parameters $\beta_k$ ($k = 4, 5, 6$) correspond to female/male earnings differentials at each of the educational levels.
Null hypotheses

\[ H_6: \beta_4 = 0, \]
\[ H_7: \beta_5 = 0, \]
\[ H_8: \beta_6 = 0, \]
\[ H_9: \beta_4 = \beta_5 = \beta_6 = 0, \]

remain interesting. Finally, the null hypothesis that the starting earnings is the same for all candidates is

\[ H_{12}: \beta_1 = \beta_2 = \beta_3 \text{ and } \beta_4 = \beta_5 = \beta_6 = 0. \]  \tag{9.7.60}
Example 9.7.7: Consider the regression model

\[ y_t = \beta_1 + \beta_2 x_t + \beta_3 z_t + u_t \quad (t = 1, 2, \ldots, T), \]  

(9.7.61)

where \( y_t \) and \( x_t \) are real aggregate personal consumption and disposable income, and

\[ z_t = \begin{cases} 
1, & \text{if year } t \text{ is a war year} \\
0, & \text{otherwise} 
\end{cases} \]  

(9.7.62)
Note that this model can also be expressed as

\[ y_t = \beta_1 + \beta_2 x_t + u_t \quad \text{(peacetime)}, \quad (9.7.63) \]

\[ y_t = (\beta_1 + \beta_3) + \beta_2 x_t + u_t \quad \text{(wartime)}. \quad (9.7.64) \]

Equations (9.7.63) and (9.7.64) imply the marginal propensity to consume is the same in peacetime and wartime, but that there is a shift \( \beta_3 \) in consumption level.

- An obvious null hypothesis of interest is

\[ H_{13}: \beta_3 = 0. \quad (9.7.65) \]
• Ordering the observations so that the \textit{first m correspond to war years},

\[
\frac{\partial \text{SSE}}{\partial \beta_1} = 2 \sum_{t=1}^{T} \hat{u}_t = 0,
\]

(9.7.66)

\[
\frac{\partial \text{SSE}}{\partial \beta_2} = 2 \sum_{t=1}^{T} x_t \hat{u}_t = 0,
\]

(9.7.67)

\[
\frac{\partial \text{SSE}}{\partial \beta_3} = 2 \sum_{t=1}^{m} \hat{u}_t = 0,
\]

(9.7.68)

Letting \( p \) denote \textit{peacetime} and \( w \) denote \textit{wartime}, it is easy to see that

\[
\bar{y}_w = \bar{y}_w \quad \text{and} \quad \bar{y}_p = \bar{y}_p.
\]

(9.7.69)
Example 9.7.8: As an alternative to Example 9.7.7, consider

\[ y_t = \beta_1 + \beta_2 x_t + \beta_3 (x_t z_t) + u_t \quad (t = 1, 2, \ldots, T), \quad (9.7.70) \]

or equivalently,

\[ y_t = \beta_1 + \beta_2 x_t + u_t \quad \text{(peacetime)}, \quad (9.7.71) \]
\[ y_t = \beta_1 + (\beta_2 + \beta_3) x_t + u_t \quad \text{(wartime)}. \quad (9.7.72) \]

Equations (9.7.71) and (9.7.72) imply the mpc increases by \( \beta_3 \) from peacetime to wartime, but that the intercept remains the same. An obvious null hypothesis of interest is again (9.7.65), but where now the wartime shift refers to the mpc.
**Example 9.7.9:** Corresponding to both Examples 9.7.7 and 9.7.8, consider

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 z_t + \beta_4 (x_t z_t) + u_t \quad (t = 1, 2, \ldots, T), \quad (9.7.73)$$

or equivalently,

$$y_t = \beta_1 + \beta_2 x_t + u_t \quad \text{(peacetime),} \quad (9.7.74)$$

$$y_t = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) x_t + u_t \quad \text{(wartime).} \quad (9.7.75)$$

In the case of (9.7.74) and (9.7.75) the intercept and the slope of consumption can shift between peacetime and wartime.
• An obvious hypothesis of interest is

\[ H_{14}: \beta_3 = \beta_4 = 0. \]  \hspace{1cm} (9.7.76)

• One implication of Example 9.7.9 is that the observations from peacetime years have no effect on the regression coefficient estimates for wartime, and vice-versa (see Exercise 9.7.6).
The peacetime and wartime models are still tied together, however, in the sense if the errors are assumed to have the same variance in the two models, then all the observations should be pooled together to estimate \( \sigma^2 \) efficiently, i.e.,

\[
s^2 = \frac{1}{T-4} \sum_{t=1}^{T} \hat{u}_t^2. \tag{9.7.77}
\]

In this case Example 9.7.9 is a special case of Example 9.7.5.

Exercise 9.7.1 provides a good test of your understanding of Section 9.7.
Exercise 9.7.1: Suppose you have observations on $T = 100$ Canadians for the following regression:

$$y = \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{10} z_{10} + X\gamma + u,$$  \hspace{1cm} (9.7.78)

where $X$ is a $T \times (K-10)$ matrix of relevant socio-economic-demographic variables, and
\[ y = \ln \text{wage rate}, \]
\[ z_1 = \text{intercept}, \]
\[ z_2 = 1, \text{if female}; = 0, \text{if male}, \]
\[ z_3 = 1, \text{if French-speaking}; = 0, \text{otherwise}, \]
\[ z_4 = 1, \text{if lives in Ontario}; = 0, \text{otherwise}, \]
\[ z_5 = 1, \text{if lives in Quebec}; = 0, \text{otherwise}, \]
\[ z_6 = z_2 z_3, \]
\[ z_7 = z_2 z_4, \]
\[ z_8 = z_2 z_5, \]
\[ z_9 = z_3 z_4, \]
\[ z_{10} = z_3 z_5. \]
Who is the excluded group (i.e., the intercept term)?
State clearly and explicitly the null and two-sided alternative hypotheses for the following tests. State whether a t-test or a F-test is appropriate and corresponding degrees of freedom.
(a) Test whether \textit{(ceteris paribus)} females have the same wage rates as males.
(b) Test whether (ceteris paribus) wage rates are the same in all regions of Canada.
(c) Test whether \textit{(ceteris paribus)} French-speaking females in Quebec have the same wage rates as French-speaking males in Ontario.
(d) Test whether \((ceteris paribus)\) a non-French-speaking male living outside of Ontario and Quebec has the same wage rate as a non-French-speaking female living outside of Ontario and Quebec.
Example: Differences in differences (DD) measures the effect of a treatment at a given period in time. It is often used to measure the change induced by a particular treatment or event.

- The basic premise of “diffs-in-diffs” is to examine the effect of some sort of treatment by comparing the treatment group after treatment both to the treatment group before treatment and to some other control group.
• Naively, you might consider simply looking at the treatment group before and after treatment to try to deduce the effect of the treatment.

• However, a lot of other things may have been going on at the exact same time as the treatment.

• The “diffs-in-diffs” method uses a control group to subtract out other changes at the same time, assuming that these other changers were identical between the treatment and control groups.
• Suppose state A passes a bill offering tax deduction to employer providing health insurance, and in the year after the bill passed (year 2) the percentage of firms offering health insurance increased by 30% compared to the year before the bill was passed (year 1).

  ○ To estimate the impact of the bill on the percentage of firms offering health insurance, we could simply do a before and after analysis and conclude that the bill increased insurance offerings by 30%.
The problem is that there could be a trend over time for more employers to offer insurance.

It is impossible to identify if the tax deductibility or the time trend caused this increase in firm offering.

One way to identify the impact of the bill is to run a DD regression. If there is a state B that did not change the way it treated employer provided health insurance, we could use this as a control group to compare the changes between A and B between the two years.
Run the regression:

\[ y_t = \beta_1 + \beta_2 \, \tilde{x}_{t1} + \beta_3 \, \tilde{x}_{t2} + \beta_4 \, \tilde{x}_{t3} + u_t \]

where \( y_t \) is the percentage of firms offering health insurance in each state in each time period. \( \tilde{x}_{t1} \) is a time dummy, \( \tilde{x}_{t2} \) is a state dummy for state A, and \( \tilde{x}_{t3} \) is the interaction of the time dummy and the state A dummy.
Suppose the percentage of firms offering insurance in each state and time period.

<table>
<thead>
<tr>
<th>State A</th>
<th>State B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>a</td>
</tr>
<tr>
<td>Year 2</td>
<td>c</td>
</tr>
</tbody>
</table>

Consider what each coefficient in the regression represents.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>a</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>c - a</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>b - a</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>(d - b) - (c - a)</td>
</tr>
</tbody>
</table>
- $\beta_1$ is the baseline average.

- $\beta_2$ represents the time trend in the control group.

- $\beta_3$ represents the differences between the two states in year 1.

- $\beta_4$ represents the difference in the changes over time.
Assuming both states have the same health insurance trends over time, we have now controlled for a possible national time trend.

We can now identify what the true impact of the tax deductibility is on employers offering insurance.
9.8 Pretest Estimators

“Pretest estimators are used all of the time. ... In the remainder of this book, we entirely ignore the problems caused by pretesting, not because they are unimportant, but because, in practice, they are generally intractable.”

Russel Davidson and James MacKinnon
Example (AP, Jan., 1987): University of California

- “The results indicate that one extra serving of fresh fruits or vegetables may decrease the risk of stroke by as much as 40 per cent, regardless of other risk factors”

- “...the researchers looked for statistical links between a variety of dietary factors and stroke. But he said the researchers did not have a clear idea of how diet could affect this health problem.”

- “When you start with a whole body of data, you are going to find
correlations between some things.”

**Example:** Sullivana, Timmermann, and White (2001, *JoE*) consider a striking example of a data-driven discovery, namely, the presence of calendar effects in stock returns.

- There is a substantial evidence of systematic abnormal stock returns related to the day of the week, the week of the month, the month of the year, the turn of the month, holidays, and so forth.

- However, this evidence has largely been considered without accounting for the intensive search preceding it.
• STW use 100 years of daily data and a bootstrap procedure that allows them to measure the distortions in statistical inference induced by data mining.

• STW find that although nominal p-values for individual calendar rules are extremely significant, once evaluated in the context of the full universe from which such rules were drawn, calendar effects no longer remain significant.
• Observed behavior suggests researchers often use hypothesis tests to choose among data analytic procedures. These tests are used as stepping stones toward obtaining “improved” estimators, predictors or subsequent hypothesis tests.

• For example, in estimation rarely do researchers use the OLS estimator $\hat{b}$ in (9.2.5) or the RLS estimator $\hat{b}^*$ in (9.3.8). Instead the more commonly used estimator is defined by a procedure which chooses between $\hat{b}$ and $\hat{b}^*$ based on the outcome of a test of $H_1: R\beta = r$ versus $H_2: R\beta \neq r$. 
Such behavior yields the pretest estimator:

\[ \tilde{\beta}_{PT} = \begin{cases} 
  b^*, & \text{if } f < c \\
  b, & \text{if } f \geq c 
\end{cases} \quad (9.8.1) \]

\[ = b - [I_{(0,c)}(f)](b - b^*), \]

where \( c = c(\alpha) = F(1 - \alpha; J, T - K) \) and

\[ I_{(a,d)}(f) = \begin{cases} 
  1, & \text{if } a < f < d \\
  0, & \text{otherwise}
\end{cases}. \quad (9.8.2) \]
• If $H_1$ is not rejected, then the \textit{restricted} estimator $b^*$ is used, whereas if $H_1$ is rejected, then the \textit{unrestricted} estimator $b$ is used.

  ◦ From the frequentist standpoint, the appropriate way to evaluate (9.8.1) is based on the resulting sampling properties of the pretest estimator $\hat{\beta}_{PT}$. 
Before discussing the sampling properties of the pretest estimator, let’s review the properties of its constituent parts \( \hat{b} \) and \( b^* \). Consider the standard multiple linear normal regression model with fixed regressors

\[
y = X\beta + u. \tag{9.8.3}
\]

- Let \( \hat{\beta} \) be an arbitrary estimator of \( \beta \) and let \( G \) be a p.d. matrix. Consider the general quadratic loss function

\[
C(\hat{\beta}; \beta, G) \equiv (\hat{\beta} - \beta)'G(\hat{\beta} - \beta).
\]
Sampling theorists are concerned with the *risk*:

$$\rho(\hat{\beta}; \beta, G) \equiv E_{y|\lambda}[C(\hat{\beta}; \beta, G)] = \text{tr}[G\{\text{MSE}(\hat{\beta}; \beta)\}], \quad (9.8.4)$$

where $\lambda = [\beta', \sigma^{-2}]'$ and

$$\text{MSE}(\hat{\beta}; \beta) \equiv E_{y|\lambda}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'].$$
Exercise 9.8.2 provides a detailed comparison of $\hat{b}$ and $b^*$ using risk function $\rho(\hat{\beta}; \beta, G)$ in (9.8.4) for various $G$.

- As expected the risk functions of $\hat{b}$ and $b^*$ cross, and each dominates over different subsets of the parameter space.
- The RLS estimator is preferred if the constraints $R\beta = r$ are “close” to being valid; otherwise, OLS is preferred.
• The “truth search” motivation of Section 9.6 amounts to the assertion that the RLS estimator \( b^* \) has no bias.

  ◦ This in turn is equivalent to the assertion that non-centrality parameter (9.6.8) satisfies \( \omega = 0 \) [Exercise 9.8.2(a)].

  ◦ Under the alternative \( H_2: R\beta \neq r, \omega > 0 \).
Consider the more general class of pretest estimators

\[ \tilde{\beta}_{PT} = \tilde{\beta}_{PT}[c(\alpha, \omega_0)] \]

\[ = \begin{cases} 
  b^*, & \text{if } f < c(\alpha, \omega_0) \\ 
  b, & \text{if } f \geq c(\alpha, \omega_0) 
\end{cases} \]  \hspace{1cm} (9.8.5)

\[ = b - [I_{(0,c[\alpha,\omega_0]}(f)](b - b^*), \]

where \( c(\alpha, \omega_0) \) is a scalar critical value for a test of

\[ H_4: \omega \leq \omega_0 \text{ versus } H_5: \omega > \omega_0 \]  \hspace{1cm} (9.8.6)

satisfying
\[ \alpha = \int_{c(\alpha, \omega_0)}^{\infty} f_2(z | J, T-K, \omega_0) \, dz, \]  

(9.8.7)

and \( f_2(z | J, T-K, \omega_0) \) is a non-central F-density with \( J \) and \( T-K \) degrees of freedom and non-centrality parameter \( \omega_0 \).
• A rejection region of size $\alpha$ for (9.8.6) is

$$ \{ f \mid f > c(\alpha, \omega_0) \}.$$

• A “truth search” motivation [Exercise 9.8.2(a)] suggests $\omega_0 = 0$ and $c(\alpha, \omega_0) = F(1-\alpha; J, T-K)$.

• Dominance of $b^*$ over $b$ in terms of MSE matrices motivates test (9.8.6) with $\omega_0 = \frac{1}{2}$ [Exercise 9.8.2(b)].
Exercise 9.8.2(c) considers the in-sample predictive case of loss function \( C(\hat{\beta}; \beta, G) \equiv (\hat{\beta} - \beta)'G(\hat{\beta} - \beta) \) with \( G = X'X \), motivates test (9.8.6) with \( \omega_0 = J/2 \).

Exercises 9.8.2(d) and 9.8.2(e) motivate other choices of \( \omega_0 \).

In each case, the choice of significance level \( \alpha \) (usually made in a cavalier manner) plays a crucial role in determining the sampling performance of the estimator.
Note:

- The pretest estimator $\tilde{\beta}_{PT}$ equals neither the OLS nor RLS estimators.
- Derivation of the sampling distribution of $\tilde{\beta}_{PT}$ is complicated because it is a stochastic mixture of the sampling distributions of $b$ and $b^*$.  
- Even calculating the moments and risk function of $\tilde{\beta}_{PT}$ is a non-trivial task.
- The next theorem compares the sampling properties of $\tilde{\beta}_{PT}$ to both $b^*$ and $b$. 
Theorem 9.8.1: Define \( \nu_0 = c(\omega_0)/[c(\omega_0) + (T-K)/J] \).

(a) \( \text{MSE}(b) - \text{MSE}(\hat{\beta}_{PT}) \) is p.s.d. if \( \omega \leq [2(2 - \nu_0)]^{-1} \) and \( T - K > 1 \).

(b) \( \text{MSE}(\hat{\beta}_{PT}) - \text{MSE}(b^*) \) is p.s.d. if \( \omega \leq \frac{1}{2} \).

(c) \( \rho(b; \beta, X'X) > \rho(\hat{\beta}_{PT}; \beta, X'X) \) if \( \omega \leq \frac{1}{4} \).
(d) Given the p.d. matrix $G$, define $V = R_0 G Q R' [RQR]'^{-1}$. Let $\xi_S$ and $\xi_L$ denote the smallest and largest characteristic roots of $V$.

\begin{align*}
(1) \quad \rho(b; \beta, G) &> \rho(\hat{\beta}_{PT}; \beta, G) \text{ if} \\
\omega &< \frac{\text{tr}(V)}{2\xi_L(2 - \nu_0)} \text{ and } T - K \geq 2.
\end{align*}

\begin{align*}
(2) \quad \rho(b; \beta, G) &< \rho(\hat{\beta}_{PT}; \beta, G) \text{ if} \\
\omega &> \frac{\text{tr}[V/\{2\xi_S[2 - \min\{1, \nu_0(1 + [T - K - 2]/[J + 4])\}\}]]}}{2 \xi_L(2 - \nu_0)} \text{ and } T - K \geq 2.
\end{align*}
(e) Given the definitions in Part (d):

1. \[ \rho(\hat{\beta}_{PT}; \beta, G) \geq \rho(b^*; \beta, G) \text{ if } \omega \leq \text{tr}[V/2\xi_L]. \]

2. \[ \rho(\hat{\beta}_{PT}; \beta, G) \leq \rho(b^*; \beta, G) \text{ if } \omega \geq \text{tr}[V]/\{2\xi_s \text{ Prob}[\{\chi^2(J)/\chi^2(T-K)\} \geq c_0J/(T-K)]\}. \]

(f) \[ \text{Bias } (\tilde{\beta}_{PT}) \leq \text{Bias}(b^*). \]
Figure 3.12 Typical risk functions for the traditional and Stein-rule pretest estimators.
The pretest estimator $\tilde{\beta}_{PT}$ is a **discontinuous** function of the data: small change in $y$ can bring about a switch from $b^*$ to $b$. *This discontinuity implies the pretest estimator is inadmissible with respect to quadratic loss [Cohen (1965)].*

- To a conscientious frequentist, inadmissibility under a standard loss function should be disturbing.

- It seems necessary to consider discontinuous loss structures in order to rescue $\tilde{\beta}_{PT}$ from inadmissibility.
Such loss structures might reward parsimony arising when the restricted estimator is chosen and the dimensionality of the problem is reduced.

Quantifying “parsimony,” is a daunting task.

See Meeden and Arnold (1979, JASA).
• Pretest estimators are the logical outcome of “diagnostic testing” of assumptions advocated in many econometric circles.

  ⊎ Even when there is no relationship between the regressand and any of the regressors, if there are many possible regressors to search over via pre-tests, then the chances of obtaining a specification which looks “significant” [judged by standard nominal levels of significance that ignore the pretesting] is very high.

  ⊎ Unless the level of significance is a decreasing function of sample size in order that the probability of both type I and type II errors go to zero asymptotically, pretesting affects the asymptotic distribution theory.
• The resiliency to such damning results explains the label “the teflon factor.”
Example [Freedman (1983, *American Statistician*)] Suppose $y_t, x_{t,1}, \ldots, x_{t,50}$ ($t = 1, 2, \ldots, 100$) $\sim$ i.i.d. $N(0, 1)$.

**First Stage:** $y_t$ is regressed on $x_{t,1}, \ldots, x_{t,50}$.

**Second Stage:** $y_t$ is regressed on only those variables whose coefficients are “significant” at the .25 level.
This is repeated ten times.

<table>
<thead>
<tr>
<th>Rep</th>
<th>First Stage</th>
<th></th>
<th></th>
<th></th>
<th>Second Stage</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>F</td>
<td>p-val</td>
<td>#25%</td>
<td>#5%</td>
<td>R²</td>
<td>F</td>
<td>p-val</td>
</tr>
<tr>
<td>1</td>
<td>.50</td>
<td>.98</td>
<td>.53</td>
<td>15</td>
<td>1</td>
<td>.36</td>
<td>3.13</td>
<td>.00005</td>
</tr>
<tr>
<td>2</td>
<td>.46</td>
<td>.84</td>
<td>.73</td>
<td>9</td>
<td>0</td>
<td>.15</td>
<td>1.85</td>
<td>.07</td>
</tr>
<tr>
<td>3</td>
<td>.52</td>
<td>1.07</td>
<td>.40</td>
<td>16</td>
<td>4</td>
<td>.36</td>
<td>2.93</td>
<td>.0007</td>
</tr>
<tr>
<td>4</td>
<td>.45</td>
<td>.83</td>
<td>.75</td>
<td>7</td>
<td>1</td>
<td>.14</td>
<td>2.13</td>
<td>.05</td>
</tr>
<tr>
<td>5</td>
<td>.57</td>
<td>1.35</td>
<td>.15</td>
<td>17</td>
<td>2</td>
<td>.44</td>
<td>3.82</td>
<td>.00002</td>
</tr>
<tr>
<td>6</td>
<td>.46</td>
<td>.84</td>
<td>.73</td>
<td>12</td>
<td>1</td>
<td>.22</td>
<td>2.06</td>
<td>.03</td>
</tr>
<tr>
<td>7</td>
<td>.41</td>
<td>.70</td>
<td>.89</td>
<td>4</td>
<td>0</td>
<td>.12</td>
<td>3.33</td>
<td>.01</td>
</tr>
<tr>
<td>8</td>
<td>.42</td>
<td>.72</td>
<td>.88</td>
<td>12</td>
<td>1</td>
<td>.27</td>
<td>2.66</td>
<td>.004</td>
</tr>
<tr>
<td>9</td>
<td>.39</td>
<td>.64</td>
<td>.94</td>
<td>8</td>
<td>0</td>
<td>.20</td>
<td>2.90</td>
<td>.006</td>
</tr>
<tr>
<td>10</td>
<td>.63</td>
<td>1.69</td>
<td>.03</td>
<td>16</td>
<td>4</td>
<td>.48</td>
<td>4.80</td>
<td>$8 \times 10^{-6}$</td>
</tr>
</tbody>
</table>
Exercise 9.8.4: Consider the following statement.

“If there are any data imaginable that would lead to abandonment of the current regression specification, then a pretesting situation exists. Whether the actual data fall in the acceptable region is immaterial.”

Explain this statement [see Hill (1985-86, *Econometric Reviews*)].