3. DUMMY VARIABLES, NONLINEAR VARIABLES AND SPECIFICATION

[1] DUMMY VARIABLES

(1) Motivation:

- We wish to estimate effects of qualitative regressors on a dependent variable.
- Data (GPA1.WF1 or GPA1.TXT from Wooldridge's Web site):

Obs: 141

1. age	in years;	2. soph	=1 if sophomore
3. junior	=1 if junior;	4. senior	=1 if senior
5. senior5	=1 if fifth year senior;	6. male	=1 if male
7. campus	=1 if live on campus;	8. business	=1 if business major
9. engineer	=1 if engineering major		
10. colGPA	MSU (Michigan State Univ.) GPA		
11. hsGPA	high school GPA;	12. ACT	'achievement' score
13. job19	=1 if job <= 19 hours;	14. job20	=1 if job >= 20 hours
15. drive	=1 if drive to campus;	16. bike	=1 if bicycle to campus
17. walk	=1 if walk to campus;	18. voluntr	=1 if do volunteer work
19. PC	=1 of pers computer at sch		
20. greek	=1 if fraternity or sorority		
21. car	=1 if own car		
22. siblings	=1 if have siblings		
23. bgfriend	=1 if boy- or girlfriend		
24. clubs	=1 if belong to MSU club		
25. skipped	avg lectures missed per week		
26. alcohol	avg # days per week drink alcohol		
27. gradMI	=1 if Michigan high school		
28. fathcoll	=1 if father college grad		
29. mothcoll	=1 if mother college grad		

(2) Comparing means of two groups:

- Wish to estimate GPA difference in PC ownership.
- $colGPA_t = \beta_1 + \beta_2 PC_t + \varepsilon_t$.
- If t owns PC, $E(y_t) = \beta_1 + \beta_2$.

  If t does not own PC, $E(y_t) = \beta_1$.
- H_o: No PC difference:

  Do OLS on (*) and test H_o: $\beta_2 = 0$.

Dependent Variable: COLGPA
Method: Least Squares
Sample: 1 141
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.989412	0.039502	75.67792	0.0000
PC	0.169517	0.062680	2.704461	0.0077

R-squared	0.049989	Mean dependent var	3.056738
Adjusted R-squared	0.043154	S.D. dependent var	0.372310
S.E. of regression	0.364188	Akaike info criterion	0.831792
Sum squared resid	18.43601	Schwarz criterion	0.873618
Log likelihood	-56.64131	F-statistic	7.314107
Durbin-Watson stat	1.941198	Prob(F-statistic)	0.007697

(3) Same slope and different intercepts

- $colGPA_t = \beta_1 + \beta_2 PC_t + \beta_3 hsGPA_t + \varepsilon_t$.
- $E(colGPA_t | PC_t=1, hsGPA_t) = (\beta_1+\beta_2) + \beta_3 hsGPA_t$

  $E(colGPA_t | PC_t=0, hsGPA_t) = \beta_1 + \beta_3 hsGPA_t$.
- H_o: No PC difference → Do t-test for H_o: $\beta_2 = 0$.

Dependent Variable: COLGPA
Method: Least Squares
Sample: 1 141
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.381856	0.300058	4.605289	0.0000
PC	0.158558	0.057200	2.771989	0.0063
HSGPA	0.473794	0.087796	5.396512	0.0000

R-squared	0.215536	Mean dependent var	3.056738
Adjusted R-squared	0.204167	S.D. dependent var	0.372310
S.E. of regression	0.332136	Akaike info criterion	0.654504
Sum squared resid	15.22339	Schwarz criterion	0.717243
Log likelihood	-43.14250	F-statistic	18.95811
Durbin-Watson stat	1.889680	Prob(F-statistic)	0.000000

(4) Several categories.

- Four different groups in data:

  $soph_t$ = 1 if t is a sophomore; = 0, otherwise.

  $junior_t$ = 1 if t is a junior; = 0, otherwise.

  $senior_t$ = 1 if t is a senior; = 0, otherwise.

  $senior5_t$ = 1 if t is a fifth year senior; = 0, otherwise.

- $colGPA_t = \beta_1 + \beta_2 junior_t + \beta_3 senior_t + \beta_4 senior5_t + \beta_5 hsGPA_t + \varepsilon_t$.

- $E(colGPA_t | t = \text{sophomore}, hsGPA_t) = \beta_1 + \beta_5 hsGPA_t$.

  $E(colGPA_t | t = \text{junior}, hsGPA_t) = (\beta_1+\beta_2) + \beta_5 hsGPA_t$.

  $E(colGPA_t | t = \text{senior}, hsGPA_t) = (\beta_1+\beta_3) + \beta_5 hsGPA_t$.

  $E(colGPA_t | t = 5^{th} \text{ year senior}, hsGPA_t) = (\beta_1+\beta_4) + \beta_5 hsGPA_t$.

- H_o: No school-year effects.

  $\rightarrow H_o: \beta_2 = \beta_3 = \beta_4 = 0$.

  $\rightarrow$ Use F-test.

- In general, if one qualitative variable has p categories, we need (p-1) dummy variables.

- Why not p dummy variables?

  - $y_t = \beta_1 + \beta_2 soph_t + \beta_3 junior_t + \beta_4 senior_t + \beta_5 senior5_t$

    $+ \beta_6 hsGPA_t + \varepsilon_t$.

  $\rightarrow$ 1st column of X = 2nd+3rd+4^{th}+5^{th} columns of X.

(5) Model with multiple dummy variables

Dependent Variable: COLGPA
Method: Least Squares
Sample: 1 141
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.397243	0.398931	3.502465	0.0006
HSGPA	0.488646	0.091262	5.354311	0.0000
PC	0.173897	0.059880	2.904118	0.0043
JUNIOR	-0.163828	0.207184	-0.790741	0.4305
SENIOR	-0.155040	0.205510	-0.754418	0.4520
SENIOR5	-0.111336	0.228068	-0.488169	0.6263
BUSINESS	0.058662	0.081985	0.715520	0.4756
ENGINEER	-0.312198	0.175188	-1.782072	0.0771
ALCOHOL	0.012776	0.021460	0.595370	0.5526
FATHCOLL	0.036843	0.062347	0.590928	0.5556
MOTHCOLL	-0.004513	0.060650	-0.074408	0.9408

R-squared	0.254073	Mean dependent var	3.056738
Adjusted R-squared	0.196694	S.D. dependent var	0.372310
S.E. of regression	0.333692	Akaike info criterion	0.717606
Sum squared resid	14.47554	Schwarz criterion	0.947651
Log likelihood	-39.59120	F-statistic	4.427974
Durbin-Watson stat	2.024310	Prob(F-statistic)	0.000023

(6) Changing intercept and slopes

- The relation between colGPA and hsGPA may be different depending on PC ownership.
- $colGPA_t = \beta_1 + \beta_2 hsGPA_t + \beta_3 PC_t + \beta_4(PC_t hsGPA_t) + \varepsilon_t$.
- $E(colGPA_t|PC_t=1,hsGPA_t) = (\beta_1+\beta_3) + (\beta_2+\beta_4)hsGPA_t$

  $E(colGPA_t|PC_t=0,hsGPA_t) = \beta_1 + \beta_2 hsGPA_t$.
- H_o: intercepts are the same → H_o: $\beta_3 = 0$.

  H_o: slopes are the same → H_o: $\beta_4 = 0$.

  H_o: both intercepts and slopes are the same → H_o: $\beta_3 = \beta_4 = 0$.

Dependent Variable: COLGPA
Method: Least Squares
Sample: 1 141
Included observations: 141

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.520780	0.402217	3.780994	0.0002
HSGPA	0.432849	0.118066	3.666157	0.0004
PC	-0.155454	0.606123	-0.256473	0.7980
PC*HSGPA	0.092199	0.177170	0.520401	0.6036

R-squared	0.217083	Mean dependent var	3.056738
Adjusted R-squared	0.199939	S.D. dependent var	0.372310
S.E. of regression	0.333017	Akaike info criterion	0.666713
Sum squared resid	15.19336	Schwarz criterion	0.750366
Log likelihood	-43.00328	F-statistic	12.66223
Durbin-Watson stat	1.880463	Prob(F-statistic)	0.000000

Wald Test:
Equation: Untitled

Null Hypothesis:	C(3)=0 C(4)=0		
F-statistic	3.957070	Probability	0.021344
Chi-square	7.914141	Probability	0.019119

[2] INTRINSICALLY LINEAR MODELS

- AInstrinsically linear model≅ is a model that is linear in β=s, but not linear in regressors and/or dependent variables.

(1) Log-linear model:

$\ln(h_t) = \beta_1 + \beta_2\ln(z_{t2}) + \beta_3\ln(z_{t3}) + \varepsilon_t$

$\rightarrow y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \varepsilon_t$.

(2) Semilog model:

$\ln(h_t) = \beta_1 + \beta_2 t + \varepsilon_t$: $M\ln(h_t)/Mt = \beta_2$ = growth rate of h_t.

$\ln(GNP_t) = 0.7 + 0.08t$: growth rate of GNP = 0.08 (8% over time).

(3) Model with interaction terms:

$y_t = \beta_1 + \beta_2 z_{t2} + \beta_3 z_{t3} + \beta_4(z_{t2}z_{t3}) + \varepsilon_t$

$\rightarrow y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t$.

$$\rightarrow \frac{\partial E(y_t \mid z_{t2}, z_{t3})}{\partial z_{t2}} = \beta_2 + \beta_4 z_{t3} .$$

$\rightarrow$ How can we estimate $\partial E(y_t x_t)/\partial z_{t2}$ at $z_{t3} = a$?

$\rightarrow$ Set R = (0,1,0,a).

$$\rightarrow R\hat{\beta} = \hat{\beta}_2 + a\hat{\beta}_4 \rightarrow se(R\hat{\beta}) = \sqrt{RCov(\hat{\beta})R'} .$$

(4) Model with nonlinear coefficients

- CES production function: $Q_t = \gamma[\delta K^{-\rho} + (1-\delta)L^{-\rho}]^{-\upsilon/\rho}\exp(\varepsilon_t)$.
- RTS = $\dfrac{\partial \ln Q}{\partial \ln K} + \dfrac{\partial \ln Q}{\partial \ln L} = \upsilon$.
- Elasticity of substitution = $\dfrac{d\ln(K^*/L^*)}{d\ln(P_L/P_K)} = \dfrac{1}{1+\rho}$.
- The production function becomes Cobb-Douglas as $\rho \to 0$.
- (A) $\ln(Q_t) = \ln(\gamma) - \upsilon/\rho\ln[\delta K_t^{-\rho} + (1-\delta)L_t^{-\rho}] + \varepsilon_t$
- (B) $\ln(Q_t)\ .\ \ln(\gamma) + \upsilon\delta\ln(K_t) + \upsilon(1-\delta)\ln(L_t)$

  $+ \rho\upsilon\delta(1-\delta)\{-(1/2)[\ln(K_t/L_t)]^2\} + \varepsilon_t$ (See Greene.)

→ $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t$, (*)

where, $y_t = \ln(Q_t)$, $x_{t2} = \ln(K_t)$, $x_{t3} = \ln(L_t)$,

$x_{t4} = -(1/2)[\ln(K_t/L_t)]^2$,

$\beta_1 = \ln(\gamma)$, $\beta_2 = \upsilon\delta$, $\beta_3 = \upsilon(1-\delta)$, $\beta_4 = \rho\upsilon\delta(1-\delta)$.

→ $\gamma = \exp(\beta_1)$; $\delta = \beta_2/(\beta_2+\beta_3)$; $\upsilon = \beta_2+\beta_3$; $\rho = \beta_4(\beta_2+\beta_3)/(\beta_2\beta_3)$.

- Estimation Procedure for models with nonlinear coefficients.

A. Delta Method:

- Let $\theta_{p\times 1} = w(\beta)$ ($p \le k$).

Example: $$\theta = \begin{pmatrix} \gamma \\ \delta \\ v \\ \rho \end{pmatrix}; \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}; w(\beta) = \begin{pmatrix} e^{\beta_1} \\ \dfrac{\beta_2}{\beta_2 + \beta_3} \\ \beta_2 + \beta_3 \\ \dfrac{\beta_4(\beta_2 + \beta_3)}{\beta_2\beta_3} \end{pmatrix}$$

- $\hat{\theta} = w(\hat{\beta})$.

- $Cov(\hat{\theta})$. $W(\hat{\beta})Cov(\hat{\beta})W(\hat{\beta})'$, where $W(\beta) = \partial w(\beta)/\partial\beta'$.

Example: $$\frac{\partial w(\beta)}{\partial \beta'} = \begin{pmatrix} e^{\beta_1} & 0 & 0 & 0 \\ 0 & \dfrac{\beta_3}{(\beta_2 + \beta_3)^2} & -\dfrac{\beta_2}{(\beta_2 + \beta_3)^2} & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -\dfrac{\beta_4}{\beta_2{}^2} & -\dfrac{\beta_4}{\beta_3{}^2} & \dfrac{(\beta_2 + \beta_3)}{\beta_2\beta_3} \end{pmatrix}.$$

B. Minimum-Distance Method (Chamberlain, 1984, Handbook)

- Let $\beta = g(\theta_{p\times 1})$ ($p \le k$).

  Example: $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}; g(\theta) = \begin{pmatrix} \ln(\gamma) \\ v\delta \\ v(1-\delta) \\ \rho v\delta(1-\delta) \end{pmatrix}$.

- Find a minimizer $\hat{\theta}_{MD}$ of $(\hat{\beta} - g(\theta))'[Cov(\hat{\beta})]^{-1}(\hat{\beta} - g(\theta))$.

- $Cov(\hat{\theta}_{MD})$. $\{G(\hat{\theta})[Cov(\hat{\beta})]^{-1}G(\hat{\theta})'\}^{-1}$, where $G(\theta) = \partial g(\theta)/\partial\theta'$.

- Facts:
  - If $p = k$, $\hat{\theta} - \hat{\theta}_{MD} \to_p 0$.
  - If $p < k$, $\hat{\theta}_{MD}$ is more efficient than $\hat{\theta}$.

A: CES PRODUCTION FUNCTION

Data: Table7_1.wf1.

K = Capital; L = Labor; Q = Valueadd

C(1) = log(γ); C(2) = υ; C(3) = -ρ; C(4) = δ;

RTS = C(2); σ = 1/(1-C(3)).

Dependent Variable: LOG(VALUEADD)
Method: Least Squares
Sample: 1 27
Included observations: 27
Convergence achieved after 20 iterations
LOG(VALUEADD)=C(1)+C(2)/C(3)*LOG(C(4)*CAPITAL^C(3)+(1-C(4))* *LABOR^C(3))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.393811	0.349742	3.985260	0.0006
C(2)	0.989675	0.063425	15.60391	0.0000
C(3)	1.415962	1.350972	1.048106	0.3055
C(4)	0.056912	0.126400	0.450256	0.6567

R-squared	0.946781	Mean dependent var	7.443631
Adjusted R-squared	0.939840	S.D. dependent var	0.761153
S.E. of regression	0.186693	Akaike info criterion	-0.38275
Sum squared resid	0.801646	Schwarz criterion	-0.19078
Log likelihood	9.167155	Durbin-Watson stat	1.933311

B: APPROXIMATED CES FUNCTION

$C(1) = \log(\gamma)$; $C(2) = \upsilon$; $C(3) = \rho$; $C(4) = \delta$

RTS = C(2); $\sigma = 1/(1+C(3))$.

Dependent Variable: LOG(VALUEADD)
Method: Least Squares
Sample: 1 27
Included observations: 27
Convergence achieved after 8 iterations
LOG(VALUEADD)=C(1)+C(2)*C(4)*LOG(CAPITAL)+C(2)*(1-C(4))
*LOG(LABOR)+C(3)*C(2)*C(4)*(1-C(4))*(-.5)
*(LOG(CAPITAL/LABOR)^2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.467726	0.408231	3.595331	0.0015
C(2)	0.988724	0.062588	15.79734	0.0000
C(4)	-0.112773	0.419436	-0.268868	0.7904
C(3)	2.454156	8.086063	0.303504	0.7642

R-squared	0.946771	Mean dependent var	7.443631
Adjusted R-squared	0.939828	S.D. dependent var	0.761153
S.E. of regression	0.186711	Akaike info criterion	-0.38256
Sum squared resid	0.801802	Schwarz criterion	-0.19058
Log likelihood	9.164515	Durbin-Watson stat	1.932465

[3] SPECIFICATION ANALYSIS

(1) Selection of Variables

- Motivation: What regressors should I use (for a given y)?

- Maximize $\bar{R}^2 = 1 - \{(T-1)/(T-k)\}(1-R^2)$.

  → $\bar{R}^2$ does not necessarily increase with k.

  → When T is large, $\bar{R}^2$. R^2.

- Minimize Amemiya's prediction criterion:

  $$PC = s^2 \{1 + (k/T)\}$$

  Or minimize $\{(T+k)/(T-k)\}(1-R^2)$.

- Minimize Akaike's information criterion:

  $$AIC = \ln(\hat{\sigma}^2) + 2k/T,\ \hat{\sigma}^2 = SSE/T.$$

- Minimize Schwarz Criterion: $[k\ln(T)]/T + \ln(\hat{\sigma}^2)$.

- Choose variables with significant t-statistics

  (stepwise regression)

(2) Testing for nonnested models

H_o: y = Xβ + error (1)

H_a: y = Zγ + error (2)

Example:

H_o: lwage = $\beta_1 + \beta_2$educ + $\beta_3$exper + $\beta_4$exper^2 + error (Mincerian)

H_a: lwage = $\gamma_1 + \gamma_2$iq + $\gamma_3$feduc + $\gamma_4$meduc

+ $\gamma_5$exper + $\gamma_6$exper^2 + ε (Somebody hating educ.)

1) The test based on an encompassing model

- Set the following general model:

$$y = \overline{X}\overline{\beta} + \overline{Z}\overline{\gamma} + W\delta + error, \quad (3)$$

where $\overline{X}$ is the set of variables in X that are not in Z, $\overline{Z}$ is defined similarly, and W is the set of variables that are in both X and Z.

→Reject H_a if $\overline{\gamma} = 0$, and reject H_o if $\overline{\beta} = 0$.

- Two problems in this approach:

a) What if $\{\overline{\beta} \neq 0 \text{ and } \overline{\gamma} \neq 0\}$ or $\{\overline{\beta} = 0 \text{ and } \overline{\gamma} = 0\}$?

b) The encompassing model may be too big.

2) J Test: Davidson and Mackinnon (1981, Econometrica)

- Construct the following auxiliary model:

$$y_t = (1-\alpha)X\beta + \alpha Z\gamma + \varepsilon. \quad (4)$$

- If H_o is correct, $\alpha = 0$. Can indirectly test H_o by testing H_oN: $\alpha = 0$.
- Let $\hat{\gamma}$ be OLS estimator from (2): and set

$$y = X\beta_* + \alpha\left(Z\hat{\gamma}\right) + error, \quad (5)$$

where $\beta_* = (1-\alpha)\beta$.

- Do OLS on (5) and estimate β_* and α jointly.
- Do a t-test for H_oN: $\alpha = 0$.

3) Cox test

See Greene (pp. 155-159).

Example of the J test:

Mincerian (H_o) against Somebody Hating Educ (SHE, H_a).

Step 1: Do regression on the wage equation of SHE.

Dependent Variable: LWAGE
Sample(adjusted): 1 932
Included observations: 722

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.566653	0.140528	39.61231	0.0000
IQ	0.007887	0.001076	7.333015	0.0000
FEDUC	0.015409	0.005570	2.766607	0.0058
MEDUC	0.012485	0.006390	1.953939	0.0511
EXPER	0.007927	0.015355	0.516223	0.6059
EXPER^2	0.000308	0.000655	0.469989	0.6385
R-squared	0.140209	Mean dependent var		6.799923

To get fitted values, type in the Eviews window: genr fity2 = lwage – resid

Step 2:

Dependent Variable: LWAGE
Sample(adjusted): 1 932
Included observations: 722

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.644662	0.677296	2.428277	0.0154
EDUC	0.050569	0.008534	5.925566	0.0000
EXPER	0.002983	0.014999	0.198846	0.8424
EXPER^2	0.000453	0.000641	0.707025	0.4798
FITY2	**0.641796**	**0.109282**	**5.872840**	**0.0000**
R-squared	0.180349	Mean dependent var		6.799923

Example: SHE (H_a) against Mincerian (H_o)

Step 1: Do regression on the Mincerian wage equation.

Dependent Variable: LWAGE
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.517432	0.124819	44.20360	0.0000
EDUC	0.077987	0.006624	11.77291	0.0000
EXPER	0.016256	0.013540	1.200595	0.2302
EXPER^2	0.000152	0.000567	0.268133	0.7887
R-squared	0.130926	Mean dependent var		6.779004

To get fitted values, type in the Eviews window, "genr fity1 = lwage – resid.

Step 2:

Dependent Variable: LWAGE
Sample(adjusted): 1 932
Included observations: 722

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.639960	0.677449	2.420787	0.0157
IQ	0.004959	0.001162	4.269469	0.0000
FEDUC	0.009964	0.005519	1.805436	0.0714
MEDUC	0.008670	0.006276	1.381319	0.1676
EXPER	-0.002548	0.015107	-0.168681	0.8661
EXPER^2	0.000556	0.000641	0.866925	0.3863
FITY1	**0.648821**	**0.109614**	**5.919137**	**0.0000**
R-squared	0.180373	Mean dependent var		6.799923

(3) Omission of Relevant Variables

Theorem:

Assume that the true model is given:

(*) $y = X\beta + \varepsilon = W\delta + Z\gamma + \varepsilon$,

where $[W,Z] = X$, W and Z are $T\times k_1$ and $T\times k_2$, respectively. All SIC hold for (*). A misspecified model is given:

(**) $y = W\delta + \eta$.

Let $\hat{\delta}$ be OLS estimator of δ from (**); and s_w^2 be s^2 from (**). Then,

$E(\hat{\delta}) = \delta + (W'W)^{-1}W'Z\gamma$;

$\mathrm{plim}_{T\to\infty}\, \hat{\delta} = \delta + [\lim_{T\to\infty}T^{-1}W'W]^{-1}[\lim_{T\to\infty}T^{-1}W'Z]\gamma$;

$E(s_w^2) = \sigma^2 + \gamma'Z'M(W)Z\gamma/(T-k_1)$;

$\mathrm{plim}_{T\to\infty}\, s_w^2 = \sigma^2 + \lim_{T\to\infty}T^{-1}\gamma'Z'M(W)Z\gamma$,

where $M(W) = I_T - W(WNW)^{-1}WN$.

Proof: Do it by yourself.

Comment:

- s_w^2 is upward biased, that is, $E(s_w^2) > \sigma^2$ unless $\gamma = 0$.
- $\hat{\delta}$ and s_w^2 from (**) are unbiased and consistent if $\gamma = 0$.
- $\hat{\delta}$ from (**) are unbiased and consistent if $W'Z = 0$.

→ If $\mathrm{plim}_{T\to\infty}T^{-1}W'Z = 0$, all t or Wald tests with heteroskedasticity and/or autocorrelation corrections are ok asymptotically.

[Example]

- Data: (WAGE2.WF1 or WAGE2.TXT – from Wooldridge's website)

# of observations (T): 935

1. wage	monthly earnings
2. hours	average weekly hours
3. IQ	IQ score
4. KWW	knowledge of world work score
5. educ	years of education
6. exper	years of work experience
7. tenure	years with current employer
8. age	age in years
9. married	=1 if married
10. black	=1 if black
11. south	=1 if live in south
12. urban	=1 if live in SMSA
13. sibs	number of siblings
14. brthord	birth order
15. meduc	mother's education
16. feduc	father's education
17. lwage	natural log of wage

- Mincerian Wage Equation

Dependent Variable: LWAGE
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.517432	0.124819	44.20360	0.0000
EDUC	0.077987	0.006624	11.77291	0.0000
EXPER	0.016256	0.013540	1.200595	0.2302
EXPER^2	0.000152	0.000567	0.268133	0.7887
R-squared	0.130926	Mean dependent var		6.779004

- Mincerian + Parents' education levels

Dependent Variable: LWAGE
Sample(adjusted): 1 932
Included observations: 722

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.465087	0.139434	39.19475	0.0000
EDUC	0.066144	0.007827	8.450955	0.0000
EXPER	0.006588	0.015178	0.434033	0.6644
EXPER^2	0.000741	0.000650	1.140507	0.2545
FEDUC	0.011868	0.005567	2.132014	0.0333
MEDUC	0.011676	0.006311	1.850041	0.0647
R-squared	0.159477	Mean dependent var		6.799923

- Mincerian + parents' education level +iq

Dependent Variable: LWAGE
Sample(adjusted): 1 932
Included observations: 722

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.219784	0.149286	34.96511	0.0000
EDUC	0.050599	0.008548	5.919137	0.0000
EXPER	0.007999	0.015003	0.533175	0.5941
EXPER^2	0.000654	0.000642	1.018736	0.3087
FEDUC	0.009964	0.005519	1.805436	0.0714
MEDUC	0.008670	0.006276	1.381319	0.1676
IQ	0.004959	0.001162	4.269469	0.0000
R-squared	0.180373	Mean dependent var		6.799923

- Mincerian + iq

Dependent Variable: LWAGE
Method: Least Squares
Sample: 1 935
Included observations: 935

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.213965	0.132947	39.21823	0.0000
EDUC	0.057326	0.007388	7.758936	0.0000
EXPER	0.015714	0.013301	1.181414	0.2377
EXPER^2	0.000165	0.000557	0.295494	0.7677
IQ	0.005787	0.000980	5.903960	0.0000

R-squared	0.162323	Mean dependent var	6.779004
Adjusted R-squared	0.158720	S.D. dependent var	0.421144
S.E. of regression	0.386279	Akaike info criterion	0.940818
Sum squared resid	138.7665	Schwarz criterion	0.966704
Log likelihood	-434.8326	F-statistic	45.05323
Durbin-Watson stat	1.812399	Prob(F-statistic)	0.000000

(3) Inclusion of irrelevant variables

Theorem:

Assume that the true model is given:

(*) $y = W\delta + \varepsilon$,

where δ is $k_1 \times 1$. A misspecified model is given:

(**) $y = X\beta + \varepsilon = W\delta + Z\gamma + \eta$,

where γ is $k_2 \times 1$. Let $\tilde{\delta}$ be the OLS estimator of δ from (*); let $\hat{\delta}$ and $\hat{\gamma}$ be OLS from (**); and let s^2 be s^2 from (**). Then,

(i) $E(\hat{\delta}) = \delta$;

(ii) $\text{plim}_{T\to\infty} \hat{\delta} = \delta$;

(iii) $E(s^2) = \sigma^2$,

(iv) $\text{plim}_{T\to\infty} s^2 = \sigma^2$;

(v) $Cov(\hat{\delta})$ - $Cov(\tilde{\delta})$ is psd, and $Cov(\tilde{\delta}) = Cov(\hat{\delta})$ only if $W'Z = 0_{k_1 \times k_2}$.

Comment:

Even if we use invalid regressors, OLS estimates are unbiased and consistent, but inefficient [less accurate].

[Proof of (i) and (v)]

Lemma 1: For (**),

$$\hat{\beta} = \begin{pmatrix} \hat{\delta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} (W'M(Z)W)^{-1}W'M(Z)y \\ (Z'M(W)Z)^{-1}Z'M(W)y \end{pmatrix}.$$

Proof: Note that:

1) y = $\hat{y} + e = X\hat{\beta} + e = W\hat{\delta} + Z\hat{\gamma} + e$.

Premultiply 1) by W′M(Z):

2) W′M(Z)y = W′M(Z)W$\hat{\delta}$ + W′M(Z)Z$\hat{\gamma}$ + W′M(Z)e

= W′M(Z)W$\hat{\delta}$

[Why?: M(Z)Z = 0 and X′e = [W,Z]′e = 0.]

From 2), $\hat{\delta}$ = $[W'M(Z)W]^{-1}W'M(Z)y$.

Similarly, $\hat{\gamma}$ = $[Z'M(W)Z]^{-1}Z'M(W)y$.

Lemma 2:

Let A and B are conformable positive definite matrices.

If A - B is positive semidefinite, $B^{-1} - A^{-1}$ is also positive semidefinite.

Proof of (i) and (v):

(i) $\hat{\delta} = [W'M(Z)W]^{-1}W'M(Z)y = [W'M(Z)W]^{-1}W'M(Z)(W\delta + \varepsilon)$

$= \delta + [W'M(Z)W]^{-1}W'M(Z)\varepsilon$

$\rightarrow E(\hat{\delta}) = \delta$.

(v) $Cov(\hat{\delta}) = \sigma^2[W'M(Z)W]^{-1}$ (Why?); and $Cov(\tilde{\delta}) = \sigma^2(W'W)^{-1}$.

$Cov(\hat{\delta}) - Cov(\tilde{\delta})$ is positive semidefinite, because $[Cov(\tilde{\delta})]^{-1}$ - $[Cov(\hat{\delta})]^{-1}$ is positive semidefinite.

[Complete the proof by yourself.]