P218 Problem Set 3

Will Hotten

Code for this problem set can be found at the following link.

Question 1

First we want to show that:

$$(y_n, c) \xrightarrow{d} (y, c)$$

Consider the function $g(y) = f(y, c)$ which is bounded and continuous and the fact that $y_n \xrightarrow{d} y$:

$$\begin{aligned}
E[g(y_n)] &\xrightarrow{d} E[g(y)] && \text{By the Portmanteu theorem and given properties of g} \\
E[f(y_n, c)] &\xrightarrow{d} E[f(y, c)] && \text{Simply rewriting} \\
(y_n, c) &\xrightarrow{d} (y, c) && \text{By the Portmanteu thm: } y_n \xrightarrow{d} y \text{ equiv. to } E[f(y_n)] \xrightarrow{d} E[f(y)]
\end{aligned}$$

Secondly we want to show that:

$$|(y_n, x_n) - (y_n, c)| \xrightarrow{p} 0$$

Simply (as we know $x_n \xrightarrow{p} 0$):

$$\begin{aligned}
|(y_n, x_n) - (y_n, c)| &= |x_n - c| \\
&\xrightarrow{p} 0
\end{aligned}$$

So we have:

a) $(y_n, c) \xrightarrow{d} (y, c)$

b) $|(y_n, x_n) - (y_n, c)| \xrightarrow{p} 0$

Remains to be shown that a) + b) implies that $(y_n, x_n) \xrightarrow{d} (y, c)$

Consider bounded continuous function, $|f(.)| \leq M$, where: $\exists K > 0, \forall (y_n, x_n), (y_n, c)$:

$$|f(y_n, x_n) - f(y_n, c)| \leq K|(y_n, x_n) - (y_n, c)|$$

In this case:

$$|E[f(y_n, x_n)] - E[f(y_n, c)]| \leq E[|f(y_n, x_n)] - f(y_n, c)|]$$

Rewriting using indicator function:

$$= E[|f(y_n, x_n)] - f(y_n, c)|\mathbb{1}|(y_n, x_n) - (y_n, c)| < \varepsilon]$$
$$+ E[|f(y_n, x_n)] - f(y_n, c)|\mathbb{1}|(y_n, x_n) - (y_n, c)| \geq \varepsilon]$$

As $|f(y_n, x_n) - f(y_n, c)| \leq K|(y_n, x_n) - (y_n, c)|$, and $|f(.)| \leq M$:

$$\leq E[K|(y_n, x_n) - (y_n, c)|\mathbb{1}|(y_n, x_n) - (y_n, c)| < \varepsilon]$$
$$+ E[2M\mathbb{1}|(y_n, x_n) - (y_n, c)| \geq \varepsilon]$$

Using the fact $E[\mathbb{1}(event)] = Pr(event)$ and $|(y_n, x_n) - (y_n, c)| < \varepsilon$ in first case:

$$\leq K\varepsilon Pr(|(y_n, x_n) - (y_n, c)| < \varepsilon)$$
$$+ 2MPr(|(y_n, x_n) - (y_n, c)| \geq \varepsilon)$$

As $Pr(|(y_n, x_n) - (y, c)| < \varepsilon) \leq 1$:

$$\leq K\varepsilon + 2MPr(|(y_n, x_n) - (y_n, c)| \geq \varepsilon)$$

$$\implies |E[f(y_n, x_n)] - E[f(y_n, c)]| \leq K\varepsilon + 2MPr(|(y_n, x_n) - (y_n, c)| \geq \varepsilon)$$

Also, using this result:

$$\begin{aligned} |E[f(y_n, x_n)] - E[f(y, c)]| &\leq |E[f(y_n, x_n)] - E[f(y_n, c)]| + |E[f(y_n, c)] - E[f(y, c)]| \\ &\leq K\varepsilon + 2MPr(|(y_n, x_n) - (y_n, c)| + |E[f(y_n, c)] - E[f(y, c)]| \end{aligned}$$

Now taking limits as n $\to \infty$

As $(y_n, c) \xrightarrow{d} (y, c)$ from earlier (a) and the Portmanteu theorem means this also holds for $E[f(.)]$ (as f(.) bounded and continuous):

$$|E[f(y_n, c)] - E[f(y, c)]| \xrightarrow{p} 0$$

As $|(y_n, x_n) - (y_n, c)| \xrightarrow{p} 0$ from earlier (b):

$$2MPr(|(y_n, x_n) - (y_n, c)| \xrightarrow{p} 0$$

$$\implies \lim_{n\to\infty} |E[f(y_n, x_n)] - E[f(y, c)]| \leq K\varepsilon$$

Furthermore, ε is just some arbitrary value

$$\implies \lim_{n\to\infty} |E[f(y_n, x_n)] - E[f(y, c)]| = 0$$

As a result:

$$E[f(y_n, x_n)] \xrightarrow{d} E[f(y, c)]$$

By the Portmanteu theorem (as f(.) bounded and continuous):

$$(y_n, x_n) \xrightarrow{d} f(y, c)$$

Question 2

2.a.

Table 1: Results

	Dependent variable:
	logTC
logQ	0.720*** (0.017)
logPL	0.436 (0.291)
logPK	−0.220 (0.339)
logPF	0.427*** (0.100)
Constant	−3.527** (1.774)
Observations	145
R^2	0.926
Adjusted R^2	0.924
Residual Std. Error	0.392 (df = 140)
F Statistic	437.686*** (df = 4; 140)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2: Results

	Dependent variable:
	logTC
logQ	0.115* (0.061)
logQsq	0.054*** (0.005)
logPL	−0.011 (0.226)
logPK	−0.578** (0.261)
logPF	0.484*** (0.077)
Constant	−0.103 (1.395)
Observations	145
R^2	0.957
Adjusted R^2	0.956
Residual Std. Error	0.299 (df = 139)
F Statistic	622.839*** (df = 5; 139)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see from table 2 that we reject the null hypothesis that the coefficient on the $\log Q^2$ variable is equal to zero, even at the 1 percent significance level.

2.b.

I select an appropriate range for β_7 of between 4 and 8.5 as this roughly corresponds to the first and ninth deciles of the values of Q_i.

2.c.

Process to generate GLS estimates is contained within the R code in the GitHub repository linked at the top of this document.

The GLS regression ran corresponds to:

$$\log(TC_i) - \log(PF_i) = \beta_1 + \beta_2 log(Q_i) + \beta_3[log(PL_i) - log(PF_i)] + \beta_4[log(PK_i) - log(PF_i)] + \beta_6 z_i$$

Where $\beta_5 = 1 - \beta_3 - \beta_4$ as a result of imposing the specified restriction.

Regression outputs for this regression are as below:

Table 3: Results

	Dependent variable:
	logTC_sub_logPF
logQ	0.433*** (0.035)
logPL_sub_logPF	0.479*** (0.164)
logPK_sub_logPF	0.046 (0.153)
z	0.228*** (0.025)
Constant	−4.043*** (0.711)
Observations	145
R^2	0.957
Adjusted R^2	0.955
Residual Std. Error	0.313 (df = 140)
F Statistic	771.958*** (df = 4; 140)
Note:	*p<0.1; **p<0.05; ***p<0.01

Meaning the GLS parameter estimates are as follows:

$$\hat{\beta}_1 = -4.043, \hat{\beta}_2 = 0.433, \hat{\beta}_3 = 0.479, \hat{\beta}_4 = 0.046, \hat{\beta}_5 = 0.475, \hat{\beta}_6 = 0.228, \hat{\beta}_7 = 6.838$$

2.d.

To calculate the standard errors we need to estimate:

$$\sigma \left(E[\frac{\partial}{\partial \theta} h(x_i, \theta_0) \frac{\partial}{\partial \theta'} h(x_i, \theta_0)]^{-1} \right)$$

Sigma can be estimated by using:

$$\hat{\sigma} = \frac{RSS_{GLS}}{n-k}$$

I wasn't sure how to compute an estimate for $E[\frac{\partial}{\partial\theta}h(x_i,\theta_0)\frac{\partial}{\partial\theta'}h(x_i,\theta_0)]^{-1}$ but once obtained the standard errors are computed by the product of these estimates scaled by $\frac{1}{n}$.

Question 3

For parts a - c see following written solution:

3.d.

Conducting the hypothesis using the test statistic developed in the question.

$$\frac{T}{\sqrt{\hat{\kappa}-1}} = 0.935$$
$$N_{0.05}(0,1) = 1.645$$

Therefore under this test procedure we fail to reject the null hypothesis that males and females have equal variance of log(wage).

Conducting the hypothesis using the standard (normal theory) test:

$$\frac{s_1^2}{s_2^2} = 1.09$$
$$F_{0.05}(n_1-1, n_2-1) = 1.084$$

Therefore under the standard test procedure we reject the null hypothesis that males and females have equal variance of log(wage).

Question 4

See following written solution:

Question 5

See following written solution:

Question 6

6.a.

Below are the outputs from the following unrestricted regression:

$$C_t = \beta_0 + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \beta_3 C_{t-3} + \beta_4 C_{t-4} + \varepsilon_t$$

Table 4: Unrestricted regression results

	Dependent variable:
	CPC
L(CPC, 1)	1.192*** (0.067)
L(CPC, 2)	−0.128 (0.106)
L(CPC, 3)	0.144 (0.105)
L(CPC, 4)	−0.207*** (0.068)
Constant	0.00002 (0.00002)
Observations	216
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.0001 (df = 211)
F Statistic	209,445.100*** (df = 4; 211)
Note:	*p<0.1; **p<0.05; ***p<0.01

Below are the outputs from the following restricted regression:

$$C_t = \beta_0 + \beta_1 C_{t-1} + \eta_t$$

Table 5: Restricted regression results

	Dependent variable:
	CPC
L(CPC, 1)	1.002*** (0.001)
Constant	0.00004** (0.00002)
Observations	219
R^2	1.000
Adjusted R^2	1.000
Residual Std. Error	0.0001 (df = 217)
F Statistic	795,298.200*** (df = 1; 217)
Note:	*p<0.1; **p<0.05; ***p<0.01

Hypothesis test of

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

can be performed through calculating test statistic

$$F = \frac{(RSS_r - RSS_u)/3}{RSS_u/(220-5)} = 8.90$$

Corresponding critical value is $F_{crit} = F_{0.95}(3, 220-5) = 2.65$

$F > F_{crit}$ therefore we can reject the null hypothesis that

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

at the 5% significant level.

6.b.

We have the issue of endogeneity here. Omitted factors in ε_t will jointly affect both period by period income and consumption growth. For example, we could expect changes to the interest rate to affect both consumption and income growth (affects both intertemporal consumption decisions and returns to saving). Endogeneity will result in the inconsistency of our standard OLS estimates.

6.c.

1st stage:

$$log(Y_t - Y_{t-1}) = \alpha_0 + \alpha_1 log(C_{t-2} - C_{t-3}) + \alpha_2 log(C_{t-3} - C_{t-4}) + \alpha_3 log(C_{t-4} - C_{t-5}) + \alpha_4 log(C_{t-5} - C_{t-6}) + \varepsilon_t$$

Or, rewritten:

$$Y = \alpha_0 + \alpha_1 w1 + \alpha_2 w2 + \alpha_3 w3 + \alpha_4 w4 + \varepsilon_t$$

Table 6: 1st stage results

	Y
w1	-0.019 (0.136)
w2	0.337** (0.137)
w3	0.170 (0.136)
w4	-0.222 (0.135)
Constant	0.004*** (0.001)
Observations	214
R^2	0.046
Adjusted R^2	0.028
Residual Std. Error	0.010 (df = 209)
F Statistic	2.509** (df = 4; 209)
Note:	*p<0.1; **p<0.05; ***p<0.01

2nd stage: Regression of fitted values from the first stage on $log(C_t - C_{t-1})$ (Y) with White standard errors computed (HC0) and presented in parentheses:

Table 7: 2nd stage results

	Y
fittedvalues	0.437** (0.203)
Constant	0.003** (0.001)
Observations	214
R^2	0.032
Adjusted R^2	0.028
Residual Std. Error	0.005 (df = 212)
F Statistic	7.087*** (df = 1; 212)
Note:	*p<0.1; **p<0.05; ***p<0.01

The estimated causal effect is that a 1% increase in income growth will lead to a 0.437% increase in consumption growth in the same period.

6.d.

Using the ”ivreg” command to perform Two-Stage Least-Squares regression with diagnostics in R, the following test p-values are computed:

Table 8: Tests for endogeneity and weak instruments

	p-value:
Weak instruments	0.0430^{*}
Wu-Hausman	0.1732
Sargan	0.0517
Note:	*p<0.1; **p<0.05; ***p<0.01

The Wu-Hausman test results in p-value which is not significant, therefore we fail to reject the null hypothesis of exogeneity of $\log(Y_t/Y_{t-1})$. Therefore there is little evidence to suggest that the OLS estimates are not BLUE (and 2SLS is not).

The Sargan test results in p-value which is not significant, therefore we fail to reject the null hypothesis of instrument exogeneity.

As a result, it would seem that it is not sensible to use the 2SLS approach as there is a lack of evidence to suggest that our independent variable is endogenous. That being said, the Sargan test does provide evidence to suggest that the instruments selected are valid.