Econometrics HW3

Michael B. Nattinger*

April 1, 2021

1 Question 20.1

For $X = 3, Y = -1 + 2X + 5X - 5 - 3X + 6 + e$. Then, the marginal effect of X on Y for $X = 3$ is $2 + 5 - 3 = 4$.

2 Question 20.3

For $m_k(x)$ to be concave then the slopes in all regions must be weakly decreasing:

$$\beta_1 \geq \beta_1 + \beta_2 \geq \beta_1 + \beta_2 + \beta_3 \geq \beta_1 + \beta_2 + \beta_3 + \beta_4$$
$$\Rightarrow \beta_2 \leq 0, \beta_3 \leq 0, \beta_4 \leq 0.$$

3 Question 20.11

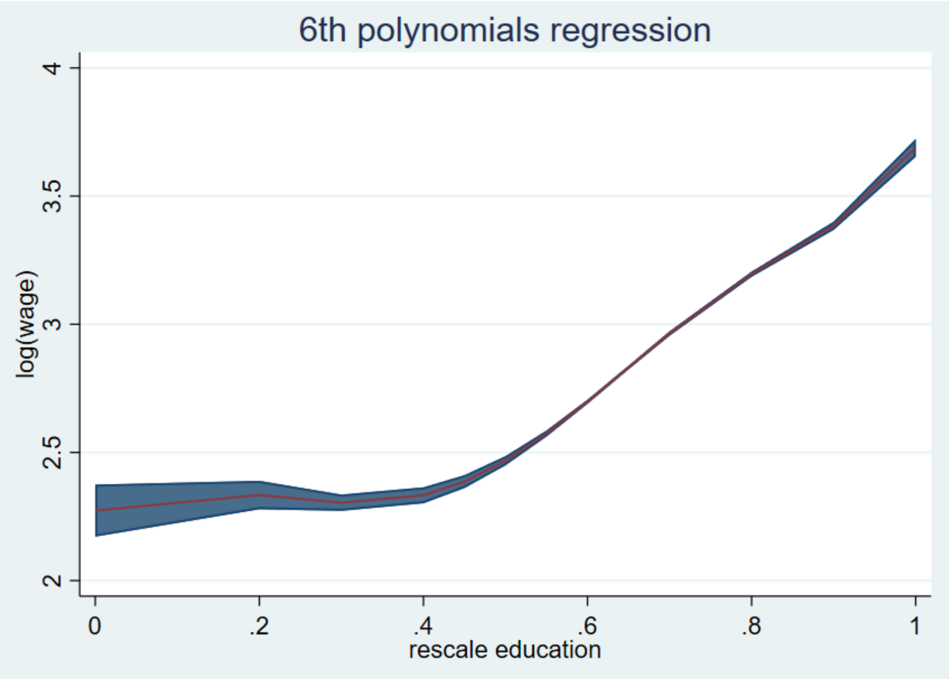
3.1 Part A

```
Linear regression                               Number of obs     =     50,742
                                                F(6, 50735)       =    2020.33
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2010
                                                Root MSE          =      .6042
```

lwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.048957	1.623172	0.65	0.518	-2.132478	4.230393
educ2	-2.096353	13.79958	-0.15	0.879	-29.14368	24.95097
educ3	-22.87369	46.22853	-0.49	0.621	-113.4821	67.73474
educ4	94.6518	74.50562	1.27	0.204	-51.38002	240.6836
educ5	-113.8697	58.00641	-1.96	0.050	-227.5629	-.1765457
educ6	44.55438	17.52917	2.54	0.011	10.19703	78.91174
_cons	2.273056	.0517622	43.91	0.000	2.171601	2.37451

*I worked on this assignment with my study group: Alex von Hafften, Andrew Smith, and Ryan Mather. I have also discussed problem(s) with Emily Case, Sarah Bass, Katherine Kwok, and Danny Edgel.

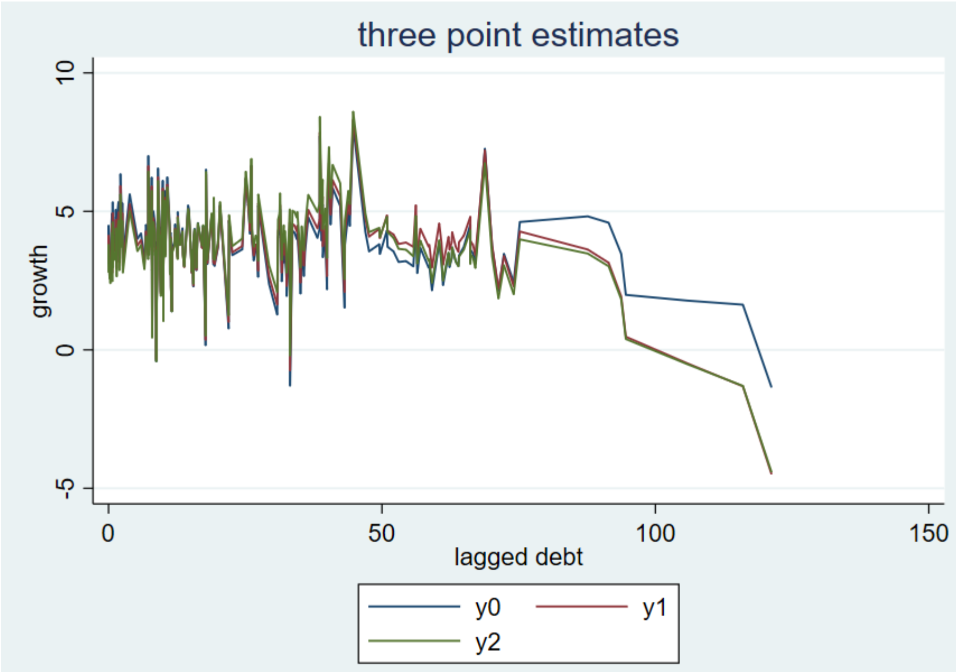
3.2 Part B



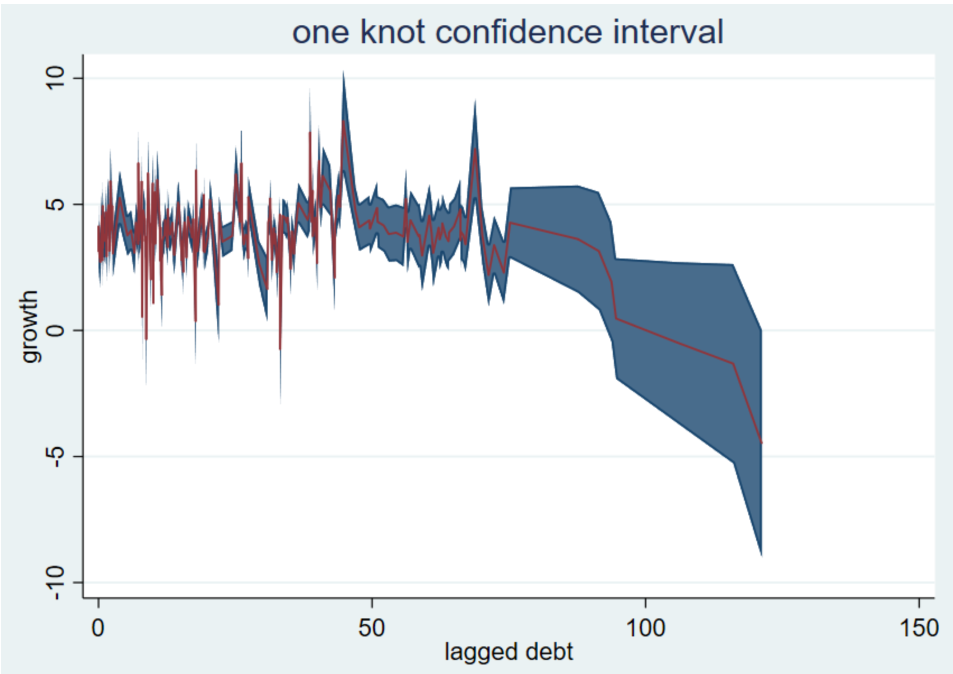

```
use "C:\Users\micha\OneDrive\Documents\HOMEWORK\Y1S1\Metrics\Q4\PS3\cps09mar.dta", clear

gen lwage = log(earnings/(hours*week))
egen educmin = min(education)
egen educmax = max(education)
gen educ = (education - educmin)/(educmax - educmin)
gen educ2 = educ^2
gen educ3 = educ^3
gen educ4 = educ^4
gen educ5 = educ^5
gen educ6 = educ^6
reg lwage educ educ2-educ6, r
predict yhat, xb
predict stdy, stdp
gen y_ub = yhat + 1.96*stdy
gen y_lb = yhat - 1.96*stdy

twoway rarea y_ub y_lb educ, sort xtitle("rescale education") ytitle("log(wage)") title(
"6th polynomials regression") legend(off)|| line yhat educ, sort legend(off)
```

4 Question 20.15

4.1 Part A



4.2 Part B



4.3 Part C

The no-spline regression has an AIC of 1249.774, one-spline has an AIC of 1248.711, and two-spline has an AIC of 1248.688. The two-spline specification has the smallest AIC, so it has the preferred specification. One-spline is also preferred to no splines, and is very close to two-splines.

4.4 Part D

From the first figure we can see that there is a substantial drop off in growth for the observations with lagged debt above 90. This is consistent in all three models, and the effect is stronger in the two models with splines in the debt level. First, note that the two models with the more substantial drop-off have better AIC values than the model with the less substantial drop-off. Note also that this drop-off occurs at 90 for all models despite not having a spline at 90 for any of the model. The effect is just as pronounced in the model with the spline at 60 as it is with the splines at 40 and 80, despite the spline at 80 being substantially closer to 90 than the spline at 60. It should be noted, however, that the model is less precisely estimated near the drop-off point, which is evident in the second figure. In my opinion the drop off at 90 is quite substantial despite the reduced precision over this region.

```
pause on
use "C:\Users\micha\OneDrive\Documents\HOMEWORK\Y1S1\Metrics\Q4\PS3\RR2010.dta", clear
gen glag = L.gdp
gen dlag = L.debt

gen dlag60 = 0
replace dlag60 = dlag-60 if dlag>=60
gen dlag40 = 0
replace dlag40 = dlag-40 if dlag>=40
gen dlag80 = 0
replace dlag80 = dlag-80 if dlag>=80

* no spline
reg gdp glag dlag
predict yhat, xb
gen y0 = yhat
estat ic
drop yhat

reg gdp glag dlag dlag60
predict yhat, xb
predict stdy, stdp
gen y_ub = yhat + 1.96*stdy
gen y_lb = yhat - 1.96*stdy
gen y1 = yhat
estat ic
drop yhat

reg gdp glag dlag dlag40 dlag80
predict yhat, xb
estat ic
gen y2 = yhat


line y0 y1 y2 dlag, sort xtitle("lagged debt") ytitle("growth") title("three point
estimates")
pause
twoway rarea y_ub y_lb dlag, sort xtitle("lagged debt") ytitle("growth") title("one knot
confidence interval") legend(off)|| line y1 dlag, sort legend(off)
```

5 Question 21.1

The difference is simply a sign difference. For $D' = 1\{X \geq c\}$ we saw that $\bar{\theta}' = m(c+) - m(c-)$, but for $D = 1\{X \geq c\}$ we now have $\bar{\theta} = m(c-) - m(c+)$.

6 Question 21.2

The identified treatment affects are at the end points c_1, c_2. We can identify $\bar{\theta}(c_1) = m(c_1+) - m(c_1-), \bar{\theta}(c_2) = m(c_2-) - m(c_2+)$.

7 Question 21.3

$$\begin{aligned} Y &= Y_0 1\{X < c\} + Y_1 1\{X \geq c\} \\ E[Y|X = x] &= E[Y_0 1\{X < c\}|X = x] + E[Y_1 1\{X \geq c\}|X = x] \\ m(x) &= m_0(x) 1\{x < c\} + m_1(x) 1\{x \geq c\} \end{aligned}$$

8 Question 21.4

With a rectangular kernel of appropriate bandwidth $2h$, $K\left(\frac{x-c}{2h}\right) = 1\{|x-c| \leq h\}$. Then the local linear estimator objective function is the following:

$$\begin{aligned} J &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 (x_i - c) D_i - \theta D_i)^2 K\left(\frac{x_i - c}{2h}\right) \\ &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - \beta_2 (x_i - c) D_i - \theta D_i)^2 1\{|x - c| \leq h\} \\ &= \sum_{|x-c| \leq h} (y_i - \beta_0 - \beta_1 x_i - \beta_2 (x_i - c) D_i - \theta D_i)^2 \end{aligned}$$

This is the ols estimation on the appropriate subsample.

9 Question 21.6

```
. reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=45.4&povrate60<=72,r
```

Linear regression

Number of obs	=	741
F(3, 737)	=	2.27
Prob > F	=	0.0791
R-squared	=	0.0075
Root MSE	=	5.2168

mort_age59..	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
povrate60	.0711572	.0652591	1.09	0.276	-.0569587	.1992732
Tnx	.1138155	.1127369	1.01	0.313	-.1075083	.3351392
T	-1.77296	.7568411	-2.34	0.019	-3.258781	-.2871383
_cons	-.8611927	3.305208	-0.26	0.795	-7.349937	5.627552

```
. reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=52.3&povrate60<=66.1,r
```

Linear regression

Number of obs	=	420
F(3, 416)	=	1.64
Prob > F	=	0.1806
R-squared	=	0.0094
Root MSE	=	5.379

mort_age59..	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
povrate60	.1356116	.200202	0.68	0.499	-.2579222	.5291453
Tnx	.0037365	.2331029	0.02	0.987	-.4544698	.4619429
T	-1.839356	1.067362	-1.72	0.086	-3.937452	.2587402
_cons	-4.505921	10.96652	-0.41	0.681	-26.06261	17.05077

```
. reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=38.4&povrate60<=80,r

Linear regression                               Number of obs     =      1,108
                                                F(3, 1104)        =       2.03
                                                Prob > F          =     0.1076
                                                R-squared         =     0.0040
                                                Root MSE          =     6.0245

------------------------------------------------------------------------------
             |               Robust
mort_age59.. |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   povrate60 |   .0607988   .0357669     1.70   0.089    -.0093799    .1309775
         Tnx |   .0167621   .0588566     0.28   0.776    -.0987214    .1322456
           T |  -1.377375   .5997218    -2.30   0.022    -2.554098   -.2006517
       _cons |  -.2612469   1.729355    -0.15   0.880     -3.65444    3.131947
------------------------------------------------------------------------------
```

The above regression outputes show the regression results from windows of $h = 8, 4, 12$, respectively. Note that $h = n$ corresponds to a bandwidth of $n\sqrt{3}$. The resulting coefficients $\bar{\theta}$, the coefficient on the variable T, are all of similar magnitude, and most are statistically significant, so the results are fairly robust to bandwidth.

10 Question 21.8

```
. reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=45.4&povrate60<=72,r

Linear regression                               Number of obs     =        741
                                                F(3, 737)         =       0.69
                                                Prob > F          =     0.5599
                                                R-squared         =     0.0027
                                                Root MSE          =     32.146

------------------------------------------------------------------------------
             |               Robust
mort_age25.. |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   povrate60 |   .0617509   .3620006     0.17   0.865    -.6489244    .7724261
         Tnx |  -.1453974   .7389214    -0.20   0.844    -1.596039    1.305244
           T |   3.480518   4.795281     0.73   0.468     -5.93352    12.89456
       _cons |   128.9467   19.11811     6.74   0.000      91.4143    166.4792
------------------------------------------------------------------------------
```

```
. reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=52.3&povrate60<=66.1,r

Linear regression                               Number of obs     =        420
                                                F(3, 416)         =       0.42
                                                Prob > F          =     0.7402
                                                R-squared         =     0.0025
                                                Root MSE          =     30.343

------------------------------------------------------------------------------
             |               Robust
mort_age25.. |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   povrate60 |  -.0674087   .9296228    -0.07   0.942    -1.894752    1.759935
         Tnx |   .3230368    1.44729     0.22   0.823    -2.521877    3.167951
           T |   2.367814   6.249574     0.38   0.705    -9.916868     14.6525
       _cons |   136.6299   51.77751     2.64   0.009     34.85169     238.408
------------------------------------------------------------------------------
```

```
. reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=38.4&povrate60<=80,r

Linear regression                               Number of obs     =      1,108
                                                F(3, 1104)        =       1.03
                                                Prob > F          =     0.3793
                                                R-squared         =     0.0025
                                                Root MSE          =     33.735
```

mort_age25..	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
povrate60	-.2029792	.2002218	-1.01	0.311	-.5958375	.189879
Tnx	-.2385138	.4011882	-0.59	0.552	-1.025691	.5486637
T	6.720869	3.890905	1.73	0.084	-.9135347	14.35527
_cons	142.7583	9.924077	14.39	0.000	123.2862	162.2305

The above regression outputs show the same robustness tests as in Question 21.8, but for the different x variable. The estimated $\bar{\theta}$ in this case is never significant at the 5% level, for most of the regressions the estimated p values are quite high. The findings of Ludwig and Miller (2007) pass this check with flying colors.

```
pause on
use "C:\Users\micha\OneDrive\Documents\HOMEWORK\Y1S1\Metrics\Q4\PS3\LM2007.dta", clear
gen T = (povrate60>=59.2)
gen Tnx = (povrate60 - 59.2)*T
* h = 8 is baseline
reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=45.4&povrate60<=72,r
reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=52.3&povrate60<=66.1,r
reg mort_age59_related_postHS povrate60 Tnx T if povrate60>=38.4&povrate60<=80,r

reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=45.4&povrate60<=72,r
reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=52.3&povrate60<=66.1,r
reg mort_age25plus_related_postHS povrate60 Tnx T if povrate60>=38.4&povrate60<=80,r
```