Econ203B Homework #2 Anton Cheremukhin

February 5, 2006

Greene Problems
4.8
$y = X\beta + \varepsilon \qquad \hat{\beta}_0 = (X'X)^{-1} X'Y \qquad Var\left[\hat{\beta}_0\right] = s_0^2 (X'X)^{-1}$
$y = X\beta + Z\gamma + \varepsilon$ Partitioned regression: $M_Z = I - Z(Z'Z)^{-1} Z'$
$\hat{\beta}_1 = (X'M_Z X)^{-1} X'M_Z Y \qquad Var\left[\hat{\beta}_1\right] = s_1^2 (X'M_Z X)^{-1}$
$\left((X'M_Z X)^{-1}\right)^{-1} - \left((X'X)^{-1}\right)^{-1} = X'M_Z X - X'X = -X'Z(Z'Z)^{-1} Z'X < 0$
Hence, $(X'M_Z X)^{-1} > (X'X)^{-1}$.
In the true regression $s_1^2 = s_0^2 = \sigma^2$ which implies that $Var\left[\hat{\beta}_0\right] < Var\left[\hat{\beta}_1\right]$.
However, when estimating $s_1^2 = e_z' e_z/(n-k-1)$ and $s_0^2 = e'e/(n-k)$
Since $e_z' e_z = e'e - \gamma^2 z' M_X z$, for the same relationship $s_1^2 > s_0^2$ to hold we need
$n - k > (1 + \gamma^2 \frac{z' M_X z}{e_z' e_z})(n-k-1) \qquad \Leftrightarrow \qquad \frac{\gamma^2}{\frac{e_z' e_z}{n-k-1}(z' M_X z)^{-1}} > 1 \qquad \Leftrightarrow \qquad |t_z| < 1$
This implies that the relationship will depend on the t-statistic of the coefficient γ.
If $\hat{\gamma} \in [-1, 1]$, then $Var\left[\hat{\beta}_0\right] < Var\left[\hat{\beta}_1\right]$. Otherwise their order is umbiguous.
4.9
$F(k, n-k) = \frac{n-k}{k} \frac{\hat{\beta}' X' X \hat{\beta}}{e'e} = \frac{n-k}{k} \frac{\left(\beta + (X'X)^{-1} X'\varepsilon\right)' X'X \left(\beta + (X'X)^{-1} X'\varepsilon\right)}{\left(\varepsilon + X(X'X)^{-1} X'\varepsilon\right)' \left(\varepsilon + X(X'X)^{-1} X'\varepsilon\right)} = \frac{n-k}{k} \frac{\varepsilon'(I-M)\varepsilon}{\varepsilon' M \varepsilon}$
Since $M(I-M) = 0$, $E[F] = \frac{n-k}{k} E\left[\frac{\varepsilon'(I-M)\varepsilon}{1}\right] E\left[\frac{1}{\varepsilon' M \varepsilon}\right] = \frac{n-k}{k} k \sigma^2 E\left[\frac{1/\sigma^2}{\chi^2(n-k)}\right] = \frac{n-k}{n-k-2}$

Exercise 1

(a) The arithmetic mean (9.0239 here) is always greater than the geometric mean (7.8396).
(b) Fractions in the sample:

total	males	females	south	non-south	union	non-union	white	hispanic	others
1.0000	0.5412	0.4588	0.7079	0.2921	0.1798	0.8202	0.8240	0.0506	0.1255

(c) Averages over subsamples:

	total	males	females	white	hispanic	others	union	non-union
LNWAGE	2.0592	**2.1653**	1.934	**2.0881**	1.8185	1.9663	**2.2935**	2.0078
EDU	13.0187	13.0138	**13.0245**	**13.1682**	11.5185	12.6418	12.8854	**13.0479**
EX	17.8221	16.9654	**18.8327**	17.7250	17.0000	**18.791**	**20.9375**	17.1393

Standard deviations over subsamples:

	total	males	females	white	hispanic	others	union	non-union
LNWAGE	0.5272	**0.5335**	0.4911	**0.5273**	0.5196	0.4931	0.4215	**0.5341**
EDU	2.6129	**2.7628**	2.4242	2.4733	**3.9756**	2.5841	**2.6215**	2.6101
EX	12.3681	12.1136	**12.587**	12.2587	**14.4299**	12.1293	**12.5257**	12.2277

(d) Positive slope coefficients indicate a positive relationship between education and wages and experience and wages. Since experience was computed as a linear combination of education, age and a constant, including age into the regression would imply exact multicollinearity and then matrix of independent variables will not have full column rank.

	const	EDU	EX
beta	0.5941	0.0964	0.0118
st.err.	0.1244	0.0083	0.0018
t-stat	4.7751	11.6031	6.7069
p-val	0.0000	0.0000	0.0000

(e) R^2=0.2115 - measures fit as a fraction of differences in wages explained by our model.

(f) P-value of the coefficient on experience (5.0995e-011) indicates that the null of no affect is rejected at the 0.05 significance level.

(g) As time passes the increase in wages due to experience starts to decline. That's what the negative sign of β_3 means. The wage is maximized at the experience level of $-b/2a$ =32.58. There is no significant change in return to education. However because the dependance on experience is non-linear (at some level experience means you are too old to work) so the linear coefficient increased a lot.

	const	EDU	EX	EX^2
beta	0.5203	0.0898	0.0349	-0.0005
st.err.	0.1236	0.0083	0.0056	0.0001
t-stat	4.2091	10.7879	6.1847	-4.3067
p-val	0.0000	0.0000	0.0000	0.0000

(h) $(\log \alpha_M = \alpha_1), (\log \alpha_F = \alpha_1 + \alpha_2)$. The negative sign of the β_2 estimate suggests that females on average get lower wage given the same education and experience. P-value of the coefficient on the female dummy (7.7514e-011) indicates that the null of no gender difference in wages in our model is rejected at the 0.05 significance level (one can't draw conclusions about presence of discrimination from this kind of models).

	const	FE	EDU	EX	EX^2
beta	0.6007	-0.2570	0.0913	0.0360	-0.0005
st.err.	0.1195	0.0387	0.0080	0.0054	0.0001
t-stat	5.0274	-6.6409	11.4054	6.6328	-4.5196
p-val	0.0000	0.0000	0.0000	0.0000	0.0000

(i) $(\log \alpha_M = \gamma_1 = \alpha_1), (\log \alpha_F = \gamma_2 = \alpha_1 + \alpha_2)$. Now we measure the constants separately for males and females. The rest of the regression is exactly the same, because we replaced one of the variables by another which is linearly dependent. The regression is a mere projection on a hyperplane, which didn't change. That's exactly what we see. The hypothesis of $H_0 : \gamma_1 = \gamma_2$ can be tested using the $\Gamma = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \end{bmatrix}$ matrix and the corresponding t-statistic. The null is rejected (p-value coincides with (h)) at the 0.05 significance level. The intercept is the sum of MA and FE, so if including a constant again the matrix of independent variables will not have full column rank.

	MA	FE	EDU	EX	EX^2
beta	0.6007	0.3437	0.0913	0.0360	-0.0005
st.err.	0.1195	0.1218	0.0080	0.0054	0.0001
t-stat	5.0274	2.8214	11.4054	6.6328	-4.5196
p-val	0.0000	0.0076	0.0000	0.0000	0.0000

(j) Testing if returns to education are sex-dependant can be done using variables, which multiply EDU by MA and FE. The hypothesis that returns to education are the same is $H_0 : \beta_3 = \beta_4$.P-value of the corresponding t-statistic (0.011307) indicates that the null of no difference is rejected at the 0.05 significance level.

	MA	FE	EDU*MA	EDU*FE	EX	EX^2
beta	0.7896	0.0358	0.0762	0.1144	0.0369	-0.0006
st.err.	0.1402	0.1713	0.0099	0.0121	0.0054	0.0001
t-stat	5.6324	0.2088	7.6770	9.4726	6.8140	-4.6804
p-val	0.0000	0.8347	0.0000	0.0000	0.0000	0.0000

(k)

To allow that return to experience differs with education level we could introduce variables like: EDU*EX, EDU*EXSQ as well as any other reasonable non-linear combinations. This could be justified by mere Taylor expansion. Here we test the hypothesis in both ways.

	const	EDU	EX	EDU*EX	EX^2
beta	0.1840	0.1131	0.0545	-0.0012	-0.0007
st.err.	0.2193	0.0151	0.0120	0.0006	0.0001
t-stat	0.8388	7.5066	4.5587	-1.8549	-4.6793
p-val	0.4019	0.0000	0.0000	0.0642	0.0000

The null of no dependance is NOT rejected at a 0.05 significance level.

	const	EDU	EX	EX^2	EDU*EX^2
beta	0.2912	0.1044	0.0406	-0.0004	-0.000026
st.err.	0.1614	0.0106	0.0062	0.0001	0.000012
t-stat	1.8043	9.8165	6.5565	-2.5127	-2.1962
p-val	0.0717	0.0000	0.0000	0.0123	0.0285

The null of no dependance IS rejected at a 0.05 significance level. So we can state that in the second setting the return to education significantly differs by education level. We are not including both new regressors due to high multicollinearity.

(l) P-value of the t-statistic on UNIO (0.0000011279) indicates that the null of no difference is rejected at the 0.05 significance level.

	const	EDU	EX	EX^2	UNIO
beta	0.5049	0.0893	0.0331	-0.0005	0.2529
st.err.	0.1210	0.0081	0.0055	0.0001	0.0513
t-stat	4.1720	10.9600	5.9679	-4.2302	4.9253
p-val	0.0000	0.0000	0.0000	0.0000	0.0000