Problem 1
Note: Please check the appendix section for detailed results and code
b) Run IV and get estimates
Code:

```
**** Generate the binary treatment variables******
gen treat1 = 0
replace treat1 =  1 if hwage >= 5
gen treat2 = 0
replace treat2 =  1 if wage >= 7
gen treat3 = 0
replace treat3 =  1 if unemployment  >= 5
***Question b*****
***Estimate the IV regression by running mrt on the binary treatment variables,
***** control variables and instrumental variable
ivregress 2sls mtr treat1 treat2 age heduc hsiblings  (siblings= treat3)
****Alternative method is to use gmm approach*****
ivregress gmm mtr treat1 treat2 age heduc hsiblings  (siblings= treat3)
```

Empirical Results:
Table 1: IV estimation

Instrumental variables (2SLS) regression

Number of obs =	753
Wald chi2(6) =	221.26
Prob > chi2 =	0.0000
R-squared =	.
Root MSE =	.09078

mtr	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
siblings	.0271812	.0607782	0.45	0.655	-.0919418	.1463042
treat1	-.0584407	.0229095	-2.55	0.011	-.1033425	-.013539
treat2	-.0626907	.0391681	-1.60	0.109	-.1394588	.0140775
age	-.0008333	.0007985	-1.04	0.297	-.0023982	.0007317
heduc	-.0089675	.0011935	-7.51	0.000	-.0113066	-.0066283
hsiblings	-.005162	.0117717	-0.44	0.661	-.0282341	.01791
_cons	.8027829	.199089	4.03	0.000	.4125755	1.19299

Instrumented: siblings
Instruments: treat1 treat2 age heduc hsiblings treat3

Alternatively:

Instrumental variables (GMM) regression

GMM weight matrix: Robust

Number of obs =	753
Wald chi2(6) =	226.51
Prob > chi2 =	0.0000
R-squared =	.
Root MSE =	.09078

mtr	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
siblings	.0271812	.0598936	0.45	0.650	-.090208	.1445704
treat1	-.0584407	.0228697	-2.56	0.011	-.1032645	-.013617
treat2	-.0626907	.039495	-1.59	0.112	-.1400995	.0147182
age	-.0008333	.0007484	-1.11	0.266	-.0023	.0006335
heduc	-.0089675	.0012995	-6.90	0.000	-.0115145	-.0064205
hsiblings	-.005162	.0115997	-0.45	0.656	-.0278971	.017573
_cons	.8027829	.1955154	4.11	0.000	.4195796	1.185986

Instrumented: siblings
Instruments: treat1 treat2 age heduc hsiblings treat3

As we can see from table 1, the result shows that both first binary treatment (treat1) and heduc have a significant impact on mtr. Unfortunately, the result shows that the instrumental variable is insignificant to the model using IV regression.

C) Calculate the bias based on the formula derived in part a

Code:

```
*****Question c******
reg mtr treat1 treat2 treat3 age heduc hsiblings
predict e_z, residuals
mat beta = e(b)
svmat beta, names(matcol)
reg siblings treat1 treat2 treat3 age heduc hsiblings
predict e_y, residuals
mat gamma = e(b)
svmat gamma, names(matcol)
scalar alpha_hat1 = betatreat3/gammatreat3
display alpha_hat1
reg treat3 treat1 treat2 age heduc hsiblings
predict e_t, residuals
* Estimate the first covariance using the second and the first residuals
corr e_y e_z, covariance
scalar scov1 = r(cov_12)
* Estimate the second covariance using the third and the first residuals
corr e_t e_z, covariance
scalar scov2 = r(cov_12)
* Finally, divide the first covariance by the second covariance.
scalar alpha_hat2 = scov1/scov2
display alpha_hat2
```

Empirical Result:

Table 2: reg mtr treat1 treat2 treat3 age heduc hsiblings

Source	SS	df	MS		
				Number of obs =	753
				F(6, 746) =	66.30
Model	1.82334348	6	.303890579	Prob > F =	0.0000
Residual	3.41922921	746	.004583417	R-squared =	0.3478
				Adj R-squared =	0.3425
Total	5.24257268	752	.006971506	Root MSE =	.0677

mtr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treat1	-.067291	.0060121	-11.19	0.000	-.0790937	-.0554882
treat2	-.0461611	.0099636	-4.63	0.000	-.0657211	-.0266012
treat3	-.0061028	.0101769	-0.60	0.549	-.0260816	.0138761
age	-.0011407	.000311	-3.67	0.000	-.0017512	-.0005302
heduc	-.0088924	.000881	-10.09	0.000	-.010622	-.0071628
hsiblings	.0000467	.0010281	0.05	0.964	-.0019717	.002065
_cons	.8964564	.0205165	43.69	0.000	.8561794	.9367333

Table 3: reg siblings treat1 treat2 treat3 age heduc hsiblings

Source	SS	df	MS
Model	196.411427	6	32.7352378
Residual	3826.14103	746	5.12887538
Total	4022.55246	752	5.34913891

Number of obs =	753
F(6, 746) =	6.38
Prob > F =	0.0000
R-squared =	0.0488
Adj R-squared =	0.0412
Root MSE =	2.2647

siblings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treat1	-.3256005	.2011154	-1.62	0.106	-.72042	.0692189
treat2	.6081235	.3332966	1.82	0.068	-.0461875	1.262434
treat3	-.224522	.3404345	-0.66	0.510	-.8928457	.4438017
age	-.0113107	.010403	-1.09	0.277	-.0317334	.009112
heduc	.0027604	.0294719	0.09	0.925	-.0550973	.0606181
hsiblings	.1916292	.0343924	5.57	0.000	.1241117	.2591467
_cons	3.446259	.6863091	5.02	0.000	2.098932	4.793586

```
. scalar alpha_hat1 = betatreat3/gammatreat3

. display alpha_hat1
.02718123
```

Table 4: reg treat3 treat1 treat2 age heduc hsiblings

Source	SS	df	MS
Model	2.42569724	5	.485139447
Residual	44.2542496	747	.059242637
Total	46.6799469	752	.062074397

Number of obs =	753
F(5, 747) =	8.19
Prob > F =	0.0000
R-squared =	0.0520
Adj R-squared =	0.0456
Root MSE =	.2434

treat3	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
treat1	.1263408	.0211147	5.98	0.000	.0848895	.1677921
treat2	.0103424	.0358189	0.29	0.773	-.0599754	.0806602
age	-.0004761	.0011179	-0.43	0.670	-.0026707	.0017186
heduc	-.0001401	.0031675	-0.04	0.965	-.0063583	.0060781
hsiblings	-.0032171	.0036944	-0.87	0.384	-.0104698	.0040357
_cons	.8733817	.0664794	13.14	0.000	.742873	1.00389

Table 5: Estimate the covariance

	e_y	e_z
e_y	**5.08795**	
e_z	**.000997**	**.004547**

	e_t	e_z
e_t	**.058849**	
e_z	**2.4e-11**	**.004547**

```
. scalar alpha_hat2 = scov1/scov2

. display alpha_hat2
41319651
```

After estimating the three regression equations showed in table 2-4, the calculated bias is given as $bias = \alpha iv - \alpha = 41319651 - 0.02718123 = 41319650.97$.