Problem Set #1:

SUGGESTED SOLUTIONS

The results are attached below.

1. The balanced panel contains larger firms (sales 120-130% bigger than the full sample on average), which are more capital and R&D intensive (capital is about 150% higher compared with 120% for employment). The set of firms that are there for at least two periods are more like the full sample than the balanced panel i.e., the difference is not just explained by firms that are only around for a year or so.

2. There is a large difference between the random effects and the fixed effects results, so the Hausman test rejects the null that they are the same. The traditional interpretation is that firm heterogeneity is important and affects firm input choices (capital, labour, R&D). This is almost certainly true here. However, it may also reflect the fact that when we are using within variation we are increasing the measurement error-to-signal ratio and that biases the coefficient. This is suggested by the fact that the within capital coefficient is small. To further study this we could examine various differences (see for example the paper by Griliches and Hausman in the Journal of Econometrics)

3. The capital coefficient falls and employment increases. This is probably due to selection effects on small firms (the balanced panel firms are larger). To see the logic, suppose that there are two sizes of firm SMALL (low capital) and LARGE (high capital) and that LARGE firms always stay in the sample whatever their ϵ but that SMALL firms exit if $\epsilon < \epsilon$.. Inthiscase, theSMALLfirms we observe will all be quite productive whereas there will be LARGE firms with low ϵ. The capital coefficient will therefore be biased downwards. Performing the probit/inverse mills ratio/Heckman style analysis does lead to the capital coefficient increasing (suggesting selection is important). Note however that R&D stock now has a negative coefficient which is unsatisfactory.

4. We now explore alternative ways to deal with correlation between the error terms and the inputs. This question uses some of the methods from the dynamic panel literature. The question did not specifically say so, but the assumption is that ω_{it} is “transmitted”. Therefore, simply using a regression that allows for AR errors (as some of you did) does not yield consistent estimates. Stata does have dynamic panel commands (see for example xtabond). However, these commands are only for linear models, where there is no AR process in ω_{it}. To answer the question you need to either program a GMM objective is Matlab, or try the approximation I wrote in the attached code (basically, I performed a grid search over ρ, for each value I computed an IV regression and choose the ρ that sets a moment condition close to zero). The code yields point estimates but does not provide the correct standard errors.

The results for the first specification, without firm-effects, are not reasonable. The signs seem wrong (labor has a negative coefficient) and the estimates are not significant. This makes sense since the identifying assumptions are probably not reasonable. The second specification is marginally more reasonable, but also is not very good. This demonstrated a general point we made in class that these methods often yield weak results.

5. In this problem we explore Olley-Pakes type estimates. Note that the question had a couple of typos. First, in the second stage the labor an capital should have been at time $t+1$. Second, the coefficients need to be restricted.

The first set of results are similar to the results from the random effects model. This should not be surprising since we do not allow for a firm specific effect. Allowing for a more general selection model (compared to the Heckman model above) does not seem to matter much. Finally, combining both the selection and investment terms changes the results very little.

Stata Output (the code is attached below)

Question 1

Variable	Obs	Mean	Std. Dev.	Min	Max
ldsal	2971	5.673087	1.960717	-.8573495	11.6984
lemp	2971	1.259177	1.775248	-3.772261	6.732211
ldnpt	2971	4.468996	2.21652	-1.389284	11.11161
ldrst	2971	3.400962	2.028775	-4.287164	9.96614
ldrnd	2971	1.78753	2.05241	-5.313206	8.432297
ldinv	2971	2.674828	2.170476	-3.844328	8.988533

Variable	Obs	Mean	Std. Dev.	Min	Max
ldsal	856	6.914288	1.837824	1.664221	11.6984
lemp	856	2.41285	1.62233	-2.071473	6.732211
ldnpt	856	5.915943	2.057395	.805678	11.11161
ldrst	856	4.885773	1.929879	.0624519	9.96614
ldrnd	856	3.221566	1.994108	-2.710516	8.432297
ldinv	856	4.068996	2.055097	-2.075558	8.88833

Variable	Obs	Mean	Std. Dev.	Min	Max
ldsal	2440	5.917722	1.942831	-.8573495	11.6984
lemp	2440	1.505869	1.734359	-3.772261	6.732211
ldnpt	2440	4.766525	2.178756	-1.389284	11.11161
ldrst	2440	3.665835	2.006672	-2.613251	9.96614
ldrnd	2440	2.018152	2.053353	-4.305191	8.432297
ldinv	2440	2.92611	2.166951	-3.844328	8.88833

Question 2

Variable	ols1	olsbe	panelwit~n	panelran~m
lemp	.49555731	.47329262	.68539345	.59840575
	.0250789	.04243702	.0361562	.0308786
ldnpt	.4601367	.48110835	.18036809	.33498823
	.01863603	.02673265	.03395214	.02639807
ldrst	.03352985	.03153441	.09893896	.06452697
	.01457263	.02897716	.03497062	.02098493
_Iyr_78	.01372579	0	.04840397	.02661348
	.03422465	0	.01642012	.01787842
_Iyr_83	-.12496231	0	-.0106752	-.07621309
	.03500908	0	.02345314	.02205521
_Iyr_88	.16335223	0	.24394114	.19598351
	.03555015	0	.02510315	.02226829
_Id357_1	-3.2352459	-1.4220112	0	-3.2950872
	.07808787	.09845477	0	.06913018
_IyrXd35_7~1	1.1707485	0	1.1350611	1.1518075
	.10422299	0	.03319479	.04498573
_IyrXd35~3_1	2.5465556	0	2.4766059	2.5037301
	.10314901	0	.04070619	.04805083
_IyrXd3~88_1	3.4755701	0	3.420753	3.4392112
	.09447972	0	.04899239	.04822963
_cons	2.8867421	2.8384729	3.5575233	3.2078862
	.05899419	.1019691	.14016269	.08953584

legend: b/se

	---- Coefficients ----			
	(b) panelwithin	(B) panelrandom	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lemp	.6853934	.5984057	.0869877	.0188091
ldnpt	.1803681	.3349882	-.1546201	.0213516
ldrst	.098939	.064527	.034412	.0279746
_Iyr_78	.048404	.0266135	.0217905	.
_Iyr_83	-.0106752	-.0762131	.0655379	.0079761
_Iyr_88	.2439411	.1959835	.0479576	.0115884
_IyrXd35_7~1	1.135061	1.151807	-.0167464	.
_IyrXd35~3_1	2.476606	2.50373	-.0271242	.
_IyrXd3~88_1	3.420753	3.439211	-.0184582	.0086115

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(9) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 51.11
Prob>chi2 = 0.0000
(V_b-V_B is not positive definite)

Question 3

Variable	olsub1	olsub2	olsub3
lemp	.57845894		.54051009
	.0126113		.01371376
ldnpt	.37226857		.39178286
	.00928791		.00994476
ldrst	.0380415		.05335209
	.00710143		.00792734
_Iyr_78	.02318958	0	.03162073
	.02043275	0	.02173807
_Iyr_83	-.07450837	-.07637714	-.08585589
	.02174854	.0140448	.02281205
_Iyr_88	.19245645	.21219132	.18087673
	.02299267	.01498475	.02492405
_Id357_1	-3.211402	1.3184095	-3.2211317
	.08554993	.03948802	.08461134
_IyrXd35_7~1	1.2387645	-.18403814	1.1961924
	.10274938	.06192879	.10504211
_IyrXd35~3_1	2.5219553	0	2.5379074
	.10324831	0	.10171051
_IyrXd3~88_1	3.6774027	-.25408593	3.564419
	.09707842	.0540959	.1056688
S5.lemp		.74008706	
		.0185214	
S5.ldnpt		.11626662	
		.01575279	
S5.ldrst		.0414079	
		.01706916	
_cons	3.1721943	.04622572	3.083116
	.03217508	.00996356	.03493123

legend: b/se

Exit Regression

```
------------------------------------------------------------------------------
Model theprobit
------------------------------------------------------------------------------

Probit regression                                 Number of obs   =       2285
                                                  LR chi2(3)      =     174.29
                                                  Prob > chi2     =     0.0000
Log likelihood = -1331.9885                       Pseudo R2       =     0.0614

------------------------------------------------------------------------------
        life |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ldnpt |  -.1830593   .0406272    -4.51   0.000    -.2626871   -.1034314
       ldrst |   .1723608   .0256799     6.71   0.000     .1220292    .2226925
       ldinv |   .1925367   .0414413     4.65   0.000     .1113133    .2737602
       _cons |   .2716402   .0907978     2.99   0.003     .0936797    .4496006
------------------------------------------------------------------------------
Heckman Type estimates

----------------------------------------
    Variable |   heck1        heck2
-------------+--------------------------
        lemp |   .5236073
             |  .01369137
       ldnpt |  .41293696
             |  .01015182
       ldrst | -.02628933
             |   .0125249
     _Iyr_78 |   .0353103            0
             |  .02145644            0
     _Iyr_83 | -.05894269   -.07189893
             |  .02275286    .01406314
     _Iyr_88 |  .17672351    .20952735
             |  .02460095    .01495651
    _Id357_1 | -3.2278199    1.3083759
             |  .08350058    .03947292
_IyrXd35_7~1 |  1.1646915   -.18167716
             |  .10373029    .06172688
_IyrXd35~3_1 |  2.5020232            0
             |  .10046713            0
_IyrXd3~88_1 |  3.5640238   -.25136621
             |  .10427657    .05392218
     imrlife |  2.8130377    .33119299
             |  .34549128    .09993373
     S5.lemp |                .7366285
             |               .01848925
    S5.ldnpt |               .11674822
             |               .01570106
    S5.ldrst |               .03897474
             |               .01702821
       _cons |  4.4456272    .17996223
             |  .17085422    .04155738
----------------------------------------
                             legend: b/se
One line estimates of Heckman Model

----------------------------------------
    Variable |   heck1        heck2
-------------+--------------------------
ldsal        |
        lemp |  .52297518
             |  .03248961
```

ldnpt	.40886041	
	.02619935	
ldrst	.03874293	
	.02220348	
_Iyr_83	-.07775473	
	.04004094	
_Id357_1	-2.0107985	
	.1688422	
_IyrXd35~3_1	1.1538213	
	.19556362	
_cons	2.4456888	
	.11158479	
death		
ldnpt	.3325072	.36681209
	.04693188	.04817407
ldrst	-.08312096	-.08432186
	.03159316	.02946309
ldinv	-.34101818	-.3500734
	.04601661	.04820712
_cons	-1.2111905	-1.3582916
	.11233675	.11430555
athrho		
_cons	1.3014812	.82328759
	.15434222	.20561676
lnsigma		
_cons	-.61102641	-1.3374807
	.07626762	.09771766
S5_ldsal		
S5.lemp		.71229232
		.0350347
S5.ldnpt		.12588716
		.02832752
S5.ldrst		-.00220212
		.03186178
_Iyr_78		.04148786
		.0249422
_Id357_1		1.2728284
		.06627436
_IyrXd35_7~1		-.089263
		.1249061
_cons		-.26502092
		.06469387

legend: b/se

Question 4

```
_bestrho:       .5410000000000003
_bests:         .2379033532088215


i.yr              _Iyr_73-88          (naturally coded; _Iyr_73 omitted)
i.d357            _Id357_0-1          (naturally coded; _Id357_0 omitted)
i.yr*i.d357       _IyrXd35_#_#        (coded as above)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    1502
-------------+------------------------------           F(  8,  1493) =   71.36
       Model |  4402.42382     8  550.302978           Prob > F      =  0.0000
    Residual |   1021.7777  1493  .684378903           R-squared     =  0.8116
-------------+------------------------------           Adj R-squared =  0.8106
       Total |  5424.20152  1501   3.6137252           Root MSE      =  .82727

------------------------------------------------------------------------------
         qdy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         qdl |  -.4227912   3.978878    -0.11   0.915    -8.227576    7.381993
         qdk |   .9907319   3.419987     0.29   0.772    -5.717759    7.699222
         qdr |  -.0497899   1.076609    -0.05   0.963    -2.161617    2.062037
     _Iyr_78 |  (dropped)
     _Iyr_83 |  -.2649514   .8208135    -0.32   0.747    -1.875022    1.345119
     _Iyr_88 |  -.2694702   1.004326    -0.27   0.788     -2.23951     1.70057
    _Id357_1 |  -2.385831   1.861643    -1.28   0.200    -6.037545    1.265883
_IyrXd35_7~1 |  (dropped)
_IyrXd35~3_1 |   1.329092     .57752     2.30   0.022     .1962555    2.461929
_IyrXd3~88_1 |   2.470861   .5338461     4.63   0.000     1.423693    3.518029
       _cons |   2.277076    5.96648     0.38   0.703    -9.426498    13.98065
------------------------------------------------------------------------------
Instrumented:  qdl qdk qdr
Instruments:   _Iyr_78 _Iyr_83 _Iyr_88 _Id357_1 _IyrXd35_78_1 _IyrXd35_83_1
               _IyrXd35_88_1 S5.lemp S5.ldnpt S5.ldrst
------------------------------------------------------------------------------


_bestrho:       .7440000000000004
_bests:         .0028113952866988

i.yr              _Iyr_73-88          (naturally coded; _Iyr_73 omitted)
i.d357            _Id357_0-1          (naturally coded; _Id357_0 omitted)
i.yr*i.d357       _IyrXd35_#_#        (coded as above)

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =     682
-------------+------------------------------           F(  6,   675) =   10.18
       Model |  1402.53916     6  233.756527           Prob > F      =  0.0000
    Residual |  857.083948   675    1.269754           R-squared     =  0.6207
-------------+------------------------------           Adj R-squared =  0.6173
       Total |  2259.62311   681  3.31809561           Root MSE      =  1.1268

------------------------------------------------------------------------------
         qdy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         qdl |   .9781527   .4678063     2.09   0.037     .0596221    1.896683
         qdk |   .3700803    .372527     0.99   0.321    -.3613708    1.101531
         qdr |   -.764131   .9754398    -0.78   0.434    -2.679392     1.15113
     _Iyr_78 |  (dropped)
     _Iyr_83 |   .0697718   .2124374     0.33   0.743    -.3473458    .4868894
```

```
     _Iyr_88 |  (dropped)
    _Id357_1 |  (dropped)
_IyrXd35_7~1 |  (dropped)
_IyrXd35~3_1 |  -.5916685   .9283171    -0.64   0.524    -2.414405    1.231068
_IyrXd3~88_1 |   1.057793   1.131495     0.93   0.350     -1.16388    3.279466
       _cons |   5.771927   1.934986     2.98   0.003     1.972612    9.571243
------------------------------------------------------------------------------
Instrumented:  qdl qdk qdr
Instruments:   _Iyr_78 _Iyr_83 _Iyr_88 _Id357_1 _IyrXd35_78_1 _IyrXd35_83_1
               _IyrXd35_88_1 S5.lemp S5.ldnpt S5.ldrst
------------------------------------------------------------------------------
```

Question 5

```
Regression on year dummies and polynomial
i.yr              _Iyr_73-88          (naturally coded; _Iyr_73 omitted)
i.d357            _Id357_0-1          (naturally coded; _Id357_0 omitted)
i.yr*i.d357       _IyrXd35_#_#        (coded as above)

      Source |       SS       df       MS              Number of obs =    2971
-------------+------------------------------           F( 17,  2953) = 4641.71
       Model |  11006.0272    17  647.413366           Prob > F      =  0.0000
    Residual |  411.877023  2953  .139477488           R-squared     =  0.9639
-------------+------------------------------           Adj R-squared =  0.9637
       Total |  11417.9042  2970   3.8444122           Root MSE      =  .37347

------------------------------------------------------------------------------
       ldsal |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        lemp |   .5843907   .0132264    44.18   0.000     .5584569    .6103246
     _Iyr_78 |    .016399   .0198845     0.82   0.410    -.0225898    .0553878
     _Iyr_83 |  -.0508266   .0215471    -2.36   0.018    -.0930754   -.0085778
     _Iyr_88 |   .1692234   .0224804     7.53   0.000     .1251446    .2133023
    _Id357_1 |  -3.244875   .0834543   -38.88   0.000    -3.408509   -3.081241
_IyrXd35_7~1 |   1.208029     .10003    12.08   0.000     1.011893    1.404164
_IyrXd35~3_1 |   2.487859   .1003619    24.79   0.000     2.291073    2.684646
_IyrXd3~88_1 |   3.653301   .0948835    38.50   0.000     3.467256    3.839345
       ldnpt |   .1962728   .0390323     5.03   0.000     .1197395    .2728061
     ldnptsq |   .0079514   .0095491     0.83   0.405    -.0107722     .026675
  ldnptldrst |  -.0075433    .009234    -0.82   0.414     -.025649    .0105624
  ldnptldinv |   .0161991   .0173665     0.93   0.351    -.0178526    .0502508
       ldrst |   .0573499   .0198489     2.89   0.004     .0184309    .0962689
     ldrstsq |  -.0021448   .0036118    -0.59   0.553    -.0092268    .0049372
  ldrstldinv |   .0025408   .0094975     0.27   0.789    -.0160816    .0211631
       ldinv |   .0719829   .0372713     1.93   0.054    -.0010975    .1450633
     ldinvsq |  -.0096643   .0095051    -1.02   0.309    -.0283017    .0089731
       _cons |   3.491645    .054206    64.41   0.000      3.38536     3.59793
------------------------------------------------------------------------------

------------------------------------------------------------------------------
Model olleypakes
------------------------------------------------------------------------------

      Source |       SS       df       MS
-------------+------------------------------         Number of obs =      1502
       Model |  1120.32629     4  280.081572         R-squared     =    0.8686
    Residual |  169.553936  1497  .113262483         Adj R-squared =    0.8682
-------------+------------------------------         Root MSE      =  .3365449
       Total |  1289.88022  1501  .859347251         Res. dev.     =   986.056

------------------------------------------------------------------------------
        newy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

/b0	3.550956	.062264	57.03	0.000	3.428822	3.67309
/b2	.341865	.0093903	36.41	0.000	.3234455	.3602844
/b3	.0647707	.0096993	6.68	0.000	.045745	.0837964
/bh	2.381869	.4155862	5.73	0.000	1.566675	3.197062
/bhsq	3.580586	.5955366	6.01	0.000	2.412411	4.748761

Parameter b0 taken as constant term in model & ANOVA table

Regression on Phat and Phatsquared

Source	SS	df	MS
Model	1069.70413	4	267.426033
Residual	220.176089	1497	.147078216
Total	1289.88022	1501	.859347251

Number of obs = 1502
F(4, 1497) = 1818.26
Prob > F = 0.0000
R-squared = 0.8293
Adj R-squared = 0.8288
Root MSE = .38351

newy	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ldnpt	.3559937	.0090631	39.28	0.000	.3382161	.3737713
ldrst	-.1032362	.0193575	-5.33	0.000	-.1412068	-.0652655
lifeprob	-2.140715	.7618922	-2.81	0.005	-3.635204	-.6462251
lifeprobsq	3.357047	.6053521	5.55	0.000	2.169619	4.544475
_cons	3.611686	.2568415	14.06	0.000	3.107879	4.115494

Model opp

Source	SS	df	MS
Model	1120.38253	7	160.054647
Residual	169.497696	1494	.113452273
Total	1289.88022	1501	.859347251

Number of obs = 1502
R-squared = 0.8686
Adj R-squared = 0.8680
Root MSE = .3368268
Res. dev. = 985.5577

newy	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
/b0	3.401648	.3362122	10.12	0.000	2.74215	4.061147
/b2	.3370969	.0106986	31.51	0.000	.3161109	.3580829
/b3	.0779428	.0194349	4.01	0.000	.0398202	.1160654
/bh	1.78504	1.096509	1.63	0.104	-.3658211	3.935902
/bhsq	3.272963	.681633	4.80	0.000	1.935903	4.610022
/bP	.3201487	.7633594	0.42	0.675	-1.177221	1.817519
/bPsq	-.1069914	.4807972	-0.22	0.824	-1.050101	.8361178
/bhP	.9038449	1.621206	0.56	0.577	-2.276237	4.083927

Parameter b0 taken as constant term in model & ANOVA table

The Stata code

```
quietly{
set mem 10m

* Load Data
use GMdata

log using PS1, replace text

* QUESTION 1
noisily display("Question 1")
* Get number of periods firm is in the sample
bysort index: egen perinsamp = count(index)
iis index
tis yr
* Get means and standard deviations
noisily sum l*
noisily sum l* if perinsamp==4
* If in the sample at least 2 periods
noisily sum l* if perinsamp>=2

* QUESTION 2
noisily display("Question 2")
*Generate computer industry indicator
gen d357 = (sic==357)
* i) regressions:
* Total
xi:regress ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp==4, robust
estimates store ols1

* Between
xi:xtreg ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp==4, be
estimates store olsbe

* Within
xi:xtreg ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp==4, fe robust
estimates store panelwithin

* Random
xi:xtreg ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp==4, robust
estimates store panelrandom

noisily estimates table ols1 olsbe panelwithin panelrandom, se

* ii) Hausman Test:
noisily hausman panelwithin panelrandom


* QUESTION 3
noisily display("Question 3")

* Total
xi:regress ldsal lemp ldnpt ldrst i.yr*i.d357
estimates store olsub1

* First Difference
xi:regress S5.ldsal S5.lemp S5.ldnpt S5.ldrst i.yr*i.d357
estimates store olsub2

* At least 2 periods Total
xi:regress ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp>=2
estimates store olsub3
```

```
noisily estimates table olsub1 olsub2 olsub3, se

* ii) Survival Regression
sort index yr
by index: generate death=index[_n+1]==.
replace death = . if yr==88
generate life = 1-death
probit life ldnpt ldrst ldinv
estimates store theprobit

noisily display("Exit Regression")
noisily estimates replay theprobit
predict lifeprob

* Inverse Mills
generate imrlife=-normalden(lifeprob)/normal(lifeprob)

* At least 2 periods Total
xi:regress ldsal lemp ldnpt ldrst i.yr*i.d357 imrlife if perinsamp>=2
estimates store heck1

tsset index yr
* First Difference at least 2 periods
xi:regress S5.ldsal S5.lemp S5.ldnpt S5.ldrst i.yr*i.d357 imrlife  if
perinsamp>=2
estimates store heck2
noisily display("Heckman Type estimates")
noisily estimates table heck1 heck2, se

/* Now there is also a simpler way of doing this in Stata given the HECKMAN
command, but it is not
really in the spirit of the question */
xi:heckman ldsal lemp ldnpt ldrst i.yr*i.d357 if perinsamp>=2,
select(death=ldnpt ldrst ldinv)
estimates store heck1
xi:heckman S5.ldsal S5.lemp S5.ldnpt S5.ldrst i.yr*i.d357 if perinsamp>=2,
select(death=ldnpt ldrst ldinv)
estimates store heck2
noisily display("One line estimates of Heckman Model")
noisily estimates table heck1 heck2, se


* QUESTION 4
noisily display("Question 4")
tsset index yr

* part (i)
local bests = 100
local bestrho = 0
forvalues rho= 0.2 (0.001) .8 {
gen qdy = ldsal - `rho'*S5.ldsal
gen qdl = lemp - `rho'*S5.lemp
gen qdk = ldnpt - `rho'*S5.ldnpt
gen qdr = ldrst - `rho'*S5.ldrst

xi: ivreg qdy (qdl qdk qdr = S5.lemp S5.ldnpt S5.ldrst)i.yr*i.d357
predict resid, residuals
gen x1 = resid*S5.ldsal
summ x1
local s1 = abs(r(N)*r(mean))

if `s1' < `bests' {
local bests = `s1'
local bestrho = `rho'
```

```
}
drop qdy qdl qdk qdr resid x1
}
noisily macro list _bestrho _bests

gen qdy = ldsal - `bestrho'*S5.ldsal
gen qdl = lemp - `bestrho'*S5.lemp
gen qdk = ldnpt - `bestrho'*S5.ldnpt
gen qdr = ldrst - `bestrho'*S5.ldrst

noisily xi: ivreg qdy (qdl qdk qdr = S5.lemp S5.ldnpt S5.ldrst)i.yr*i.d357
drop qdy qdl qdk qdr


* (ii)
local bests = 100
local bestrho = 0
forvalues rho= 0.2 (0.001) .8 {
gen qdy = (ldsal - `rho'*S5.ldsal) - (S5.ldsal - `rho'*S10.ldsal)
gen qdl = (lemp - `rho'*S5.lemp) - (S5.lemp - `rho'*S10.lemp)
gen qdk = (ldnpt - `rho'*S5.ldnpt) - (S5.ldnpt - `rho'*S10.ldnpt)
gen qdr = (ldrst - `rho'*S5.ldrst) - (S5.ldrst - `rho'*S10.ldrst)

xi: ivreg qdy (qdl qdk qdr = S5.lemp S5.ldnpt S5.ldrst)i.yr*i.d357
predict resid, residuals
gen x1 = resid*S10.ldsal
summ x1
local s1 = abs(r(N)*r(mean))

if `s1' < `bests' {
local bests = `s1'
local bestrho = `rho'
}
drop qdy qdl qdk qdr resid x1
}
noisily macro list _bestrho _bests

gen qdy = (ldsal - `bestrho'*S5.ldsal) - (S5.ldsal - `bestrho'*S10.ldsal)
gen qdl = (lemp - `bestrho'*S5.lemp) - (S5.lemp - `bestrho'*S10.lemp)
gen qdk = (ldnpt - `bestrho'*S5.ldnpt) - (S5.ldnpt - `bestrho'*S10.ldnpt)
gen qdr = (ldrst - `bestrho'*S5.ldrst) - (S5.ldrst - `bestrho'*S10.ldrst)

noisily xi: ivreg qdy (qdl qdk qdr = S5.lemp S5.ldnpt S5.ldrst)i.yr*i.d357
drop qdy qdl qdk qdr



* QUESTION 5: OLLEY-PAKES
noisily display("Question 5")
* i)

gen ldnptsq = ldnpt*ldnpt
gen ldrstsq = ldrst*ldrst
gen ldinvsq = ldinv*ldinv
gen ldnptldrst=ldnpt*ldrst
gen ldnptldinv=ldnpt*ldinv
gen ldrstldinv=ldrst*ldinv
noisily display("Regression on year dummies and polynomial")
noisily xi:reg ldsal lemp i.yr*i.d357 ldnpt* ldrst* ldinv*


* Generate new dependent variable
tsset index yr
gen newy = F5.ldsal-_b[lemp]*F5.lemp-_b[_Iyr_78]*F5._Iyr_78-
```

```
_b[_Iyr_83]*F5._Iyr_83-_b[_Iyr_88]*F5._Iyr_88-_b[_Id357_1]*F5._Id357_1-
_b[_IyrXd35_78_1]*F5._IyrXd35_78_1-_b[_IyrXd35_83_1]*F5._IyrXd35_83_1-
_b[_IyrXd35_88_1]*F5._IyrXd35_88_1
gen polynew =
_b[ldnpt]*ldnpt+_b[ldnptsq]*ldnptsq+_b[ldrst]*ldrst+_b[ldrstsq]*ldrstsq+_b[ldi
nv]*ldinv+_b[ldinvsq]*ldinvsq+_b[ldnptldrst]*ldnptldrst+_b[ldnptldinv]*ldnptld
inv+_b[ldrstldinv]*ldrstldinv

* ii) Non-linear Least Squares

nl (newy = {b0}+{b2}*F5.ldnpt+{b3}*F5.ldrst+{bh}*(polynew-{b2}*ldnpt-
{b3}*ldrst)+{bhsq}*(polynew-{b2}*ldnpt-{b3}*ldrst)^2),
estimates store olleypakes
noisily estimates replay olleypakes

*iii) Propensity Scores
gen lifeprobsq = lifeprob*lifeprob

noisily display("Regression on Phat and Phatsquared")
noisily reg newy ldnpt ldrst lifeprob lifeprobsq


*iv) Mixed polynomial propensity score
nl (newy = {b0}+{b2}*F5.ldnpt+{b3}*F5.ldrst+{bh}*(polynew-{b2}*ldnpt-
{b3}*ldrst)+{bhsq}*(polynew-{b2}*ldnpt-
{b3}*ldrst)^2+{bP}*lifeprob+{bPsq}*lifeprobsq+{bhP}*lifeprob*(polynew-
{b2}*ldnpt-{b3}*ldrst))
estimates store opp
noisily estimates replay opp

}
log close
display("end")
```