ECON4135: Solution to Written Paper 2

25th October 2007

General comments

Some general comments regarding the problem set, and your answers:

- Most answers were not very good. You'll need to improve, in order to get good grades on the exam. Read through this solution set to see what was expected (some tips and comments are also included, these were not expected).
- Some of you misunderstood some of the problems. I've generally been fairly tolerant and employed considerable good will when correcting.
- You should always include output from Stata, to show what you have done (preferably logs stating both commands used, and the results you get - see some more comments in footnote 1). However, if you have much Stata-output, it may be better to append it to the paper, rather than include it in the text, as this make the paper difficult to read.
- Remember, you are *economists*, not statisticians! So, while it of course is crucial to do the estimation and calculations correctly, don't stop there. Try to give your results a (brief) economic interpretation. This is of course particularly important when you are explicitly asked to comment or interpret the results.
- Also, proofread your writing. Try to avoid nonsensical sentences, and generally try to be concise.

Problem 1

See the appended log-file for Stata-commands used and the output Stata produced.[1] [2] I have summarized the 2003 data, results for 2004 are similar. From the log we see that we have 4215 firms, but with some missing observations on the VA_empl-variable. Number of employees

[1] The appended log documents the entire Stata-session, answering all problems of this paper. When reporting results, you should always include a (part of a) log, specifying both the command you used, and the output from the program. You can do this either by inserting the output into the document, or by appending a log. When estimating large models you may exclude the irrelevant coefficients. The appended or inserted log will be a lot easier to read if you use a fixed width font (for example `courier`) or format the regression output as a table.

[2] A note on Stata syntax: All commands, and all variables in Stata, can be abbreviated, as long as they are still unambiguous. Thus, writing `sum RD` is equivalent to writing `summarize RD_subsidy`. Also, when simultaneously referring to several variables, it may not be necessary to write all their names, see `help varlist` in Stata.

ranges from 0 (possibly misreporting?) to 3378, with an average of 30. Average tax deduction (for all firms) was 59' kroner, but only 14 percent of the firms actually got a deduction.

The firms which got a deduction are, on average, larger (mean number of employees is 61), have more highly educated employees and have a higher value added, but a larger share of these firms have no payable tax. The average subsidy within those firms who did get something is 420' kroner, ranging from 2.5' kroner to 1.6M kroner (which, interestingly, is exactly twice the upper limit). The median (at 352' kroner) is smaller than the mean, and the distribution seems to be skewed to the right. This is not unexpected, given that you cannot get a subsidy smaller than 0, but some few firms will get large subsidies.

Problem 2

We want to estimate the equation

$$\begin{aligned}\ln RDsubsidy_i \quad &= \quad \beta_0 + \beta_1 \cdot taxposition_i + \beta_2 \cdot share\_high_i \\ &\quad + \beta_3 \cdot VA\_empl_i + \beta_4 \cdot firmage\_10y_i + \beta_5 \cdot emply + u_i \qquad (1)\end{aligned}$$

In order to estimate it, we make the following assumptions about the error term, u_i:

$$E(u_i|X_i) \quad = \quad 0 \qquad (2)$$

$$(RDsubsidy_i, X_i) \quad are \quad i.i.d.\ vectors \qquad (3)$$

$$var(u_i|X_i) \quad = \quad \sigma^2 \qquad (4)$$

In the equations above X_i refers to the entire vector of covariates, i.e. *taxposition*, *share_high* etc.

Assumption (2) is essential, it ensures that $cov(u_i, X_i) = 0$. This is required for the OLS-estimators to be unbiased and consistent, that is the estimators are on average correct, and as the number of observations increases the probability that the estimators will be very different from the true values becomes small.[3]

Assumption (3) assures that the error terms are independent across observations.

The last assumption is of *homoskedasticity* (i.e., equal error term variance across all observations). This assumption is *not* required for the OLS-estimators to be unbiased or consistent, but if it is not satisfied the estimated standard errors of the OLS-estimators will be misleading, and there will exist other unbiased estimators with smaller standard errors. We'll soon return to this.

Using Stata to estimate the model for 2003, we get the results shown below. You can find results for 2004 in the appended log.

```
. reg RD tax share VA firm emply

      Source |       SS       df       MS              Number of obs =    4084
-------------+------------------------------           F(  5,  4078) =   31.65
       Model |  5176058.31     5  1035211.66           Prob > F      =  0.0000
    Residual |   133374640  4078  32705.8951           R-squared     =  0.0374
-------------+------------------------------           Adj R-squared =  0.0362
```

[3]Note that as long as we include a constant term in the regression, $E(u_i) = 0$ is not restrictive. The critical part of assumption (2) is that the covariates does not contain any information about the error term.

```
       Total |   138550699  4083  33933.5534           Root MSE      =  180.85

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -25.62379    5.72527    -4.48   0.000    -36.84845   -14.39914
  share_high |   220.8335   25.34942     8.71   0.000     171.1348    270.5322
     VA_empl |   .0027261   .0037569     0.73   0.468    -.0046394    .0100916
 firmage_10y |  -8.516276   5.701123    -1.49   0.135    -19.69359    2.661037
       emply |   .1238319   .0173638     7.13   0.000     .0897894    .1578744
       _cons |   66.48483   5.233204    12.70   0.000      56.2249    76.74477
------------------------------------------------------------------------------
```

From the computer output we see that the conditional expectation is given as[4]

$$\begin{aligned} E(RDsubsidy_i|X_i) &= 66.5 - 25.6 \cdot taxposition_i + 221 \cdot share\_high_i \\ &\quad +.00273 \cdot VA\_empl_i - 8.52 \cdot firmage\_10y_i + .124 \cdot emply \end{aligned}$$

Thus, we see that there is a negative correlation between subsidy and positive payable tax, while the share of highly educated employees and number of employees both correlates positively with the subsidy. There is no significant effect[5] of value added or age of the firm.

The assumption of homoskedasticity may be overly restrictive (optimistic?), heteroskedasticity is often a problem in cross-sectional data like these. This means that the error term variance may not be constant over firms ($var(u_i|S_i, E_i) = \sigma_i^2$), for example we expect the range of potential variation to be larger for larger firms (e.g. firms with more employees). In order to handle this we can use robust standard errors, as is done in the regression output below (still using data for 2003, with results for 2004 given in the appendix):

```
. reg RD tax share VA firm emply ,robust

Linear regression                                      Number of
obs =    4084
                                                       F(  5,  4078) =   12.15
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0374
                                                       Root MSE      =  180.85

------------------------------------------------------------------------------
             |               Robust
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -25.62379   5.712823    -4.49   0.000    -36.82404   -14.42354
```

[4]When reporting regression results, try to use a meaningful level of precision. Report at least the two first non-zero digits, if you report $\beta_3 = 0.003$, this could mean anything from 0.0025 to 0.0035, which may be an important difference. However, it is seldom relevant to report more than than three to four digits either, the twelfth digit will typically neither be precisely estimated nor interesting. However, when doing calculations you should include a few extra decimal places to avoid error due to lacking numerical precision.

[5]If you are to be prudent, 'effect' is a strong word. It implies a statement about causality, which may not always be warranted.

```
  share_high |   220.8335   41.06181     5.38   0.000       140.33    301.3371
     VA_empl |   .0027261   .0053747     0.51   0.612    -.0078113    .0132635
 firmage_10y |  -8.516276   5.637936    -1.51   0.131    -19.56971    2.537156
       emply |   .1238319   .0444131     2.79   0.005     .0367579    .2109059
       _cons |   66.48483   5.617412    11.84   0.000     55.47164    77.49803
------------------------------------------------------------------------------
```

Much is unchanged: All the coefficient estimates and the R^2. The `robust`-option makes Stata calculate the estimated standard errors in a different way however, so these are different, and thus the t- and p-values also change. More specifically, the standard errors increase (except for $firmage\_10y$ and $taxposition$, which are marginally reduced), but to a different degree: While the standard errors of $share\_high$ and VA_empl increase somewhat, there is a more than twofold increase in the standard errors of *emply*.

Problem 3

In order to increase the fit of the model, we want to replace $\beta_5 \cdot emply$ in eq. (1) with the dummy set $\sum_{k=2}^{5} \gamma_k \cdot empl\{k\}$. The regression output is given below.[6]

```
. reg RD tax share VA firm empl2-empl5

      Source |       SS       df       MS              Number of obs =    4084
-------------+------------------------------           F(  8,  4075) =   66.62
       Model |  16025704.5     8  2003213.06           Prob > F      =  0.0000
    Residual |   122524994  4075  30067.4832           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1139
       Total |   138550699  4083  33933.5534           Root MSE      =   173.4

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -20.63591   5.498826    -3.75   0.000    -31.41661   -9.855205
  share_high |   229.1315   24.36033     9.41   0.000     181.3719     276.891
     VA_empl |  -.0000786   .0036064    -0.02   0.983     -.007149    .0069918
 firmage_10y |    7.31752   5.552952     1.32   0.188    -3.569299    18.20434
       empl2 |   22.09559   7.289712     3.03   0.002     7.803771    36.38741
       empl3 |   75.84837   6.724101    11.28   0.000     62.66546    89.03128
       empl4 |   187.8132   13.00368    14.44   0.000     162.3189    213.3076
       empl5 |   189.2877   13.37939    14.15   0.000     163.0568    215.5186
       _cons |   14.61535   6.425678     2.27   0.023     2.017514    27.21319
------------------------------------------------------------------------------
```

We see that this model fits better, R^2 increases from 0.037 to 0.116. Also, the similar increase in adjusted R^2 indicates that this reflects a true increase in explanative power, not just more

[6] I have, for the added output, chosen to stick to non-robust estimation in the following, even though the results in the last problem indicated we may have an issue with heteroskedasticity here. However, remember that even if this is the case, our estimates are still unbiased and consistent, but the standard errors may be misleading.

variables. Inspecting the coefficients, we see find the cause: they do not reflect a linear relationship. Rather, after a rapid initial rise in the expected subsidy with employees, the marginal effect of additional employees becomes practically zero.[7] In light of this non-linearity, I'll stick to the dummy specification for the rest of the problem set.

If we also included $empl1$ in the model, we would have the situation that $\sum_{k=1}^{5} empl\{k\} = 1 = constant\quad term$. I.e., some of the variables in the regression would be a linear combination of each other, and we would have a problem with multicollinearity. This makes it impossible to estimate the model, we cannot distinguish between the effect of the employment categories and the constant term. In order to avoid this problem, we must eliminate one variable, to make this the reference category (Stata would automatically have dropped one of the employment-dummies).[8]

Perfect multicollinearity is rarely a problem with continuous variables, these are very unlikely to be linear combinations of each other[9], but when using categorical variables it is easy to include a complete set of dummies.

There is nothing special about $empl1$, so we could just as well have excluded any of the other categories to make up the reference (or we could have excluded the constant term). This will change the estimated employment-coefficients, but the standard errors will not change, and neither will the differences between the coefficients. Thus, if we rather excluded $empl2$ (this amount to forcing $\hat{\gamma}_2 = 0$), we would get $\tilde{\gamma}_1 = -22.1$, $\tilde{\gamma}_3 = 75.8 - 22.1 = 53.7$ etc. As the coefficient changes, and the standard errors stay the same, the t- and p-values will also change. This is because the coefficients are now tested against a different null hypothesis.

Problem 4

A 99% CI is given as

$$[\hat{\beta} - t^c_{df,0.005} \cdot \hat{se}, \hat{\beta} + t^c_{df,0.005} \cdot \hat{se}], \tag{5}$$

where $t^c_{df,0.005}$ is the critical t-value for a two-sided test at the 99% level of significance, using a t-distribution with $df = N - K - 1$ (the number of observations minus the number of covariates minus one (for the constant term)) degrees of freedom. Below I show how to calculate the lower bound for the coefficient of VA_empl, using Stata's `display`-command, and the stored values from the estimation:[10]

```
. di "VA_empl, 99% ci lower bound: " %9.5g _b[VA] - invttail(e(df_r),0.005)*_se[VA]
VA_empl, 99%ci lower bound: -.0093723
```

[7] This indicates that a promising, and more parsimonious specification could be to rather use some concave function of $empl$, such as log($empl$) - try it!

[8] The problem set explicitly asks you to answer this, without reestimating the model. Then you should do just that, almost all of you have estimated the model with the extra dummy. If this is necessary for you to see what will happen, you will need to study for the exam!

[9] It is possible that such variables are highly, but not perfectly correlated, this is called imperfect multicollinearity. In such cases estimation is possible, but standard errors increase.

[10] Stata stores coefficients in the vector `_b`, standard errors in the vector `_se`, and several other useful stuff in `e(·)` - try running `ereturn list` after an estimation. Using these saves copying, marginally increases precision, and is extremely practical if you want to write programs (as opposed to using Stata interactively).

An easier way of doing this, however, is just to get the 99% CI's directly from the estimation, using the option `level(99)`:[11] See the appended log for Stata command and output.[12] Thus we directly get the relevant CI's:

```
  share_high : [166, 292]
     VA_empl : [-.00937, .00921]
 firmage_10y : [-6.99, 21.6]
```

The meaning of a 99% CI is that if we estimate a (correctly specified!) regression model on many (independent) samples, the CI would encompass the true parameter value 99% of the times.

We see that the confidence intervals for both VA_empl and $firmage\_10y$ encompass zero, thus we conclude that of these three variables, only $share\_high$ is a significant determinant of the subsidy. Firms with a higher share of employees at the highest educational level tend to get a larger subsidy. This may reflect that most R&D is done by highly educated staff, thus the expected amount of R&D done in a firm, and thus the subsidy, increases with this share.

Problem 5

From the above regression output, and the one for 2004 in the appended log, we see that 99% CI's for *taxposition* are:[13]

```
 2003 : [-34.8, -6.47]
 2004 : [-51.2, -21.5]
```

We see that although the coefficients may seem to be different, the confidence intervals do overlap, so we have no strong evidence for claiming there is a change. A test statistic to check for this could be:

$$\begin{aligned} t &= \frac{\hat{\beta}_{2003} - \hat{\beta}_{2004}}{\hat{se}(\hat{\beta}_{2003} - \hat{\beta}_{2004})} = \frac{\hat{\beta}_{2003} - \hat{\beta}_{2004}}{\sqrt{\hat{se}(\hat{\beta}_{2003})^2 + \hat{se}(\hat{\beta}_{2004})^2}} \\ &= \frac{-20.63591 - -36.38267}{\sqrt{5.498826^2 + 5.768243^2}} = 1.98, \end{aligned}$$

see the appended log for calculations.[14] This is to be compared with a critical value, for a two-sided test with 99% level of significance, the value of 2.58 from the normal distribution gives a more than sufficient approximation here. We see that $t < t^c$, and thus conclude, as we did from inspecting the CI's, that we can not reject the null hypothesis, of unchanged coefficients.

[11] Yet another option, if you're just interested in one or a few variables, and don't want to run the entire regression (this may be a hassle, if the number of observations and covariates is large) is Stata's command `lincom`. See the log-file, and look it up in Stata's help system!

[12] Some of you estimated a regression equation containing just the variables for which you need a CI. You should rather stick to the complete specification, controlling for other covariates as well.

[13] Very many of you estimated a regression equation containing just *taxposition*. You should rather stick to the complete specification, controlling for other covariates as well.

[14] Testing is a more of a hassle in this case, because I'm using results from two different regressions, and thus can't use Stata's internal commands. Thus, the test in Problem 7 is easier to perform. For this test, I've just copied the values from the regression output.

Problem 6

See the appended log for the relevant commands and Stata output. From the log we see that the estimated coefficient of the 2003 dummy is -8.40, and that this is significant at the 95% level of significance (although not at the 99% level). Thus, expected subsidies are larger in 2004 than in 2003. This may reflect several different explanations, e.g. the subsidies may be adjusted to reflect inflation, although in that case the increase may seem large (compare to the average of 58.7 found for 2003 in problem 2004). Other possible explanations may be increased funding for the scheme resulting in larger pay-outs, the firms may have increased research or just gotten better at writing applications.

If we included a dummy also for 2004, we would have that $d\_2003 + d\_2004 = 1 = constant$, i.e. multicollinearity, as in Problem 3. Thus, in order to be able to estimate the model, Stata would have dropped either of the dummies.

Problem 7

The relevant test of the relationship between the expected subsidy and *taxposition* is just the regression in the last problem. Thus, from the Stata output associated with Problem 6, we see that *taxposition* is negatively related to the subsidy (with a coefficient of -28.3) and highly significant (with a t-value of -7.12).

Interpreting this is not straight-forward. If this reflects large R&D expenditures and low income for start-up firms, it should be captured by the age-variable. We would expect that getting the subsidy as a tax cut or in cash doesn't matter to the firms. But it may be that broke firms have larger utility of liquidity, and thus get an extra incentive to write applications. Also, it may be the case that some managers/firms are just good at getting the best from the public sector, both tax exemption and R&D subsidies.

Problem 8

Assumption (3) states that the dependent variables, and also the regressors, should be i.i.d. This implies that the error terms are uncorrelated: $cov(u_i, u_j) = 0, \quad i \neq j$. This will likely not be the case when firms appear twice. It seems likely that firm characteristics are persistent over time, and that a firm with a large positive (negative) residual in 2003, will also have a positive (negative) residual in 2004. This may reflect that the firm for example has a large R&D department (compared to other firms with similar observable characteristics), which is likely to be persistent between years. If we use $u_{i,t}$ to denote the residual of firm i in year t, this implies $cov(u_{i,2003}, u_{i,2004}) \neq 0$.

This is usually referred to as *autocorrelation*, and has a impact on the estimates similar to that of heteroskedasticity: The coefficient estimates will not be affected, but the estimated standard errors will. Thus, we will likely overstate the precision of the estimates, and may reject hypotheses we shouldn't have rejected. In Stata, we can control for such error term correlations using the option `cluster(orgnr)` to `regress`.

Problem 9

The fact that only about 14 percent of the firms got a subsidy means that the subsidy cannot be anywhere near normally distributed. (The distribution of subsidies > 0 is also somewhat skewed, that is a minor problem however.) Thus, the regression model we have used so far may be inappropriate. In the appendix I have included output from a regression using instead y, a dummy for whether the firm got any subsidy, as the dependent variable.[15]

It's difficult to make meaningful comparisons of the magnitudes of the coefficients, given the different natures of y (binary) and $RD\_subsidy$ (continuous). However, comparing the current regression output with that from problem 6, we see that all variables have the same signs, and all the t-values are similar. Thus the qualitative picture stays the same.

[15] The standard approach when using a binary left hand side variable is to use either a `logit` or `probit` model. The linear model we use, often called the linear probability model, has some conceptual and practical problems, but is still consistent and often considered a good starting point.

Appendix: Stata log

```
------------------------------------------------------------------------------
      log:  \\Balder\540$\kir\Internett\Annet\ECON 4135\wp2.log
 log type:  text
opened on:  25 Oct 2007, 15:20:57

. /* Stata-code for written paper II
>  * ECON 4135 , Autumn 2007 */
. . . * Problem 1 . use manuf2003,clear

. su emply- empl5

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       emply |      4215    30.32076    160.9113          0       3378
  share_high |      4215    .0273272    .1117486          0          1
     VA_empl |      4084    397.3964    758.7823     -13000      28340
 taxposition |      4215    .5333333    .4989468          0          1
 firmage_10y |      4215     .462159    .4986252          0          1
-------------+--------------------------------------------------------
  RD_subsidy |      4215    58.72997     182.738          0       1600
           y |      4215    .1399763    .3470036          0          1
       empl2 |      4215    .2185053    .4132811          0          1
       empl3 |      4215    .2994069    .4580526          0          1
       empl4 |      4215    .0483986    .2146324          0          1
-------------+--------------------------------------------------------
       empl5 |      4215    .0455516    .2085353          0          1

. su emply- empl5 if y==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       emply |       590    60.76271    191.3244          0       3378
  share_high |       590    .0570886    .1306425          0          1
     VA_empl |       583    424.6133    638.6254      -3777      11049
 taxposition |       590    .4508475    .4980004          0          1
 firmage_10y |       590    .4457627    .4974714          0          1
-------------+--------------------------------------------------------
  RD_subsidy |       590    419.5709    295.3912      2.515       1600
           y |       590           1           0          1          1
       empl2 |       590    .1576271    .3647002          0          1
       empl3 |       590    .4389831    .4966841          0          1
       empl4 |       590    .1440678    .3514564          0          1
-------------+--------------------------------------------------------
       empl5 |       590    .1118644     .315467          0          1
```

```
. su RD_subsidy if y==1,de

                         RD_subsidy
-------------------------------------------------------------
      Percentiles      Smallest
 1%        12.84          2.515
 5%       38.122          7.564
10%       78.458          9.153       Obs                 590
25%      165.002         10.781       Sum of Wgt.         590

50%     352.2135                      Mean           419.5709
                        Largest       Std. Dev.      295.3912
75%          720           1440
90%          800           1600       Variance       87255.96
95%      822.551           1600       Skewness       .6600611
99%     1223.629           1600       Kurtosis       3.115796

. . * Problem 2 . reg RD tax share VA firm emply

      Source |       SS       df       MS              Number of obs =    4084
-------------+------------------------------           F(  5,  4078) =   31.65
       Model |  5176058.31     5  1035211.66           Prob > F      =  0.0000
    Residual |   133374640  4078  32705.8951           R-squared     =  0.0374
-------------+------------------------------           Adj R-squared =  0.0362
       Total |   138550699  4083  33933.5534           Root MSE      =  180.85

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -25.62379    5.72527    -4.48   0.000    -36.84845   -14.39914
  share_high |   220.8335   25.34942     8.71   0.000     171.1348    270.5322
     VA_empl |   .0027261   .0037569     0.73   0.468    -.0046394    .0100916
 firmage_10y |  -8.516276   5.701123    -1.49   0.135    -19.69359    2.661037
       emply |   .1238319   .0173638     7.13   0.000     .0897894    .1578744
       _cons |   66.48483   5.233204    12.70   0.000      56.2249    76.74477
------------------------------------------------------------------------------

. /* Note: We only need to use as many letters of variable names,
>  * as to make them unique and thus understandable to Stata.
>  * Also note, Stata is case-sensitive! */
. . use manuf2004

. reg RD tax share VA firm emply

      Source |       SS       df       MS              Number of obs =    4100
-------------+------------------------------           F(  5,  4094) =   50.34
       Model |  8744545.34     5  1748909.07           Prob > F      =  0.0000
```

```
    Residual |   142226879  4094  34740.3222           R-squared     =  0.0579
-------------+------------------------------           Adj R-squared =  0.0568
       Total |   150971424  4099  36831.2819           Root MSE      =  186.39

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -41.80898     6.0566    -6.90   0.000    -53.68321   -29.93475
  share_high |    280.488   26.13402    10.73   0.000     229.2511    331.7249
     VA_empl |   .0111815   .0028806     3.88   0.000     .0055339    .0168291
 firmage_10y |   6.076629   5.837064     1.04   0.298    -5.367189    17.52045
       emply |   .1288355    .017327     7.44   0.000     .0948652    .1628059
       _cons |   73.28736   5.735425    12.78   0.000     62.04281    84.53191
------------------------------------------------------------------------------

. . use manuf2003

. reg RD tax share VA firm emply ,robust

Linear regression                                      Number of
obs =    4084
                                                       F(  5,  4078) =   12.15
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0374
                                                       Root MSE      =  180.85

------------------------------------------------------------------------------
             |               Robust
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -25.62379   5.712823    -4.49   0.000    -36.82404   -14.42354
  share_high |   220.8335   41.06181     5.38   0.000       140.33    301.3371
     VA_empl |   .0027261   .0053747     0.51   0.612    -.0078113    .0132635
 firmage_10y |  -8.516276   5.637936    -1.51   0.131    -19.56971    2.537156
       emply |   .1238319   .0444131     2.79   0.005     .0367579    .2109059
       _cons |   66.48483   5.617412    11.84   0.000     55.47164    77.49803
------------------------------------------------------------------------------

. use manuf2004

. reg RD tax share VA firm emply ,robust

Linear regression                                      Number of
obs =    4100
                                                       F(  5,  4094) =   17.77
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0579
```

```
                                                       Root MSE      =  186.39

------------------------------------------------------------------------------
             |               Robust
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -41.80898   6.433852    -6.50   0.000    -54.42283   -29.19514
  share_high |    280.488   47.55889     5.90   0.000     187.2467    373.7292
     VA_empl |   .0111815   .0047637     2.35   0.019     .0018421    .0205209
 firmage_10y |   6.076629   5.819209     1.04   0.296    -5.332184    17.48544
       emply |   .1288355   .0458719     2.81   0.005     .0389017    .2187694
       _cons |   73.28736   6.080932    12.05   0.000     61.36543    85.20929
------------------------------------------------------------------------------

. . * Problem 3 . use manuf2003

. reg RD tax share VA firm empl2-empl5

      Source |       SS       df       MS              Number of obs =    4084
-------------+------------------------------           F(  8,  4075) =   66.62
       Model |  16025704.5     8  2003213.06           Prob > F      =  0.0000
    Residual |   122524994  4075  30067.4832           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1139
       Total |   138550699  4083  33933.5534           Root MSE      =   173.4

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -20.63591   5.498826    -3.75   0.000    -31.41661   -9.855205
  share_high |   229.1315   24.36033     9.41   0.000     181.3719     276.891
     VA_empl |  -.0000786   .0036064    -0.02   0.983     -.007149    .0069918
 firmage_10y |    7.31752   5.552952     1.32   0.188    -3.569299    18.20434
       empl2 |   22.09559   7.289712     3.03   0.002     7.803771    36.38741
       empl3 |   75.84837   6.724101    11.28   0.000     62.66546    89.03128
       empl4 |   187.8132   13.00368    14.44   0.000     162.3189    213.3076
       empl5 |   189.2877   13.37939    14.15   0.000     163.0568    215.5186
       _cons |   14.61535   6.425678     2.27   0.023     2.017514    27.21319
------------------------------------------------------------------------------

. . * Problem 4
. di "VA_empl, 99% ci lower bound: " %9.5g _b[VA] - invttail(e(df_r),0.005)*_se[VA]
VA_empl, 99% ci lower bound: -.0093723

. di "VA_empl, 99% ci upper bound: " %9.5g _b[VA] + invttail(e(df_r),0.0055)*_se[VA]
VA_empl, 99% ci upper bound: .0092151

. lincom VA ,level(99)
```

```
 ( 1)  VA_empl = 0

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0000786   .0036064    -0.02   0.983    -.0093723    .0092151
------------------------------------------------------------------------------

. reg RD tax share VA firm empl2-empl5 ,level(99)

      Source |       SS       df       MS              Number of obs =    4084
-------------+------------------------------           F(  8,  4075) =   66.62
       Model |  16025704.5     8  2003213.06           Prob > F      =  0.0000
    Residual |   122524994  4075  30067.4832           R-squared     =  0.1157
-------------+------------------------------           Adj R-squared =  0.1139
       Total |   138550699  4083  33933.5534           Root MSE      =   173.4

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -20.63591   5.498826    -3.75   0.000    -34.80658   -6.465234
  share_high |   229.1315   24.36033     9.41   0.000      166.354    291.9089
     VA_empl |  -.0000786   .0036064    -0.02   0.983    -.0093723    .0092151
 firmage_10y |    7.31752   5.552952     1.32   0.188    -6.992639    21.62768
       empl2 |   22.09559   7.289712     3.03   0.002     3.309736    40.88144
       empl3 |   75.84837   6.724101    11.28   0.000     58.52012    93.17662
       empl4 |   187.8132   13.00368    14.44   0.000     154.3023    221.3242
       empl5 |   189.2877   13.37939    14.15   0.000     154.8085    223.7669
       _cons |   14.61535   6.425678     2.27   0.023    -1.943853    31.17456
------------------------------------------------------------------------------

. . * Problem 5 . use manuf2004

. reg RD tax share VA firm empl2-empl5 ,level(99)

      Source |       SS       df       MS              Number of obs =    4100
-------------+------------------------------           F(  8,  4091) =   91.31
       Model |  22872312.7     8  2859039.09           Prob > F      =  0.0000
    Residual |   128099112  4091  31312.4203           R-squared     =  0.1515
-------------+------------------------------           Adj R-squared =  0.1498
       Total |   150971424  4099  36831.2819           Root MSE      =  176.95

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [99% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -36.38267   5.768243    -6.31   0.000    -51.24762   -21.51773
```

```
  share_high |   290.8909   24.86771    11.70   0.000     226.8061    354.9758
     VA_empl |   .0101209   .0027357     3.70   0.000     .0030709    .0171709
 firmage_10y |    23.0223   5.621733     4.10   0.000     8.534916    37.50969
       empl2 |   24.42863   7.418571     3.29   0.001     5.310742    43.54653
       empl3 |   79.16296   6.852952    11.55   0.000     61.50268    96.82323
       empl4 |   223.5669   13.14094    17.01   0.000     189.7023    257.4315
       empl5 |   206.6539   13.61469    15.18   0.000     171.5684    241.7394
       _cons |   14.71667   6.856328     2.15   0.032    -2.952307    32.38564
------------------------------------------------------------------------------

. di "test statistic: "
(-20.63591--36.38267)/sqrt(5.498826^2+5.768243^2) test statistic:
1.9759281

. . * Problem 6 . use manuf2003

. append using manuf2004

. gen d_2003 = year==2003

. reg RD tax share VA firm empl2-empl5 d_2003

      Source |       SS       df       MS              Number of obs =    8184
-------------+------------------------------           F(  9,  8174) =  138.68
       Model |  38364878.2     9  4262764.25           Prob > F      =  0.0000
    Residual |   251259295  8174  30738.8421           R-squared     =  0.1325
-------------+------------------------------           Adj R-squared =  0.1315
       Total |   289624173  8183  35393.3977           Root MSE      =  175.32

------------------------------------------------------------------------------
  RD_subsidy |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -28.34934   3.981638    -7.12   0.000    -36.15436   -20.54432
  share_high |   259.1364   17.40769    14.89   0.000      225.013    293.2599
     VA_empl |   .0063219   .0021743     2.91   0.004     .0020597    .0105841
 firmage_10y |    15.0352   3.953013     3.80   0.000     7.286293    22.78412
       empl2 |   23.17059   5.204471     4.45   0.000     12.96851    33.37268
       empl3 |   77.43843    4.80391    16.12   0.000     68.02155    86.85531
       empl4 |   205.3759   9.250252    22.20   0.000      187.243    223.5087
       empl5 |    198.003   9.549805    20.73   0.000     179.2829     216.723
      d_2003 |  -8.403869   3.893134    -2.16   0.031     -16.0354   -.7723367
       _cons |   18.25918   5.192032     3.52   0.000     8.081481    28.43689
------------------------------------------------------------------------------

. . * Problem 9 . reg y tax share VA firm empl2-empl5 d_2003

      Source |       SS       df       MS              Number of obs =    8184
```

```
-------------+------------------------------           F(  9,  8174) =  121.51
       Model |  123.907048     9  13.7674498           Prob > F      =  0.0000
    Residual |  926.122155  8174  .113300973           R-squared     =  0.1180
-------------+------------------------------           Adj R-squared =  0.1170
       Total |   1050.0292  8183  .128318368           Root MSE      =   .3366

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 taxposition |  -.0523249   .0076442    -6.84   0.000    -.0673096   -.0373402
  share_high |    .395625   .0334206    11.84   0.000     .3301122    .4611379
     VA_empl |   .0000112   4.17e-06     2.69   0.007     3.06e-06    .0000194
 firmage_10y |   .0258087   .0075893     3.40   0.001     .0109318    .0406857
       empl2 |   .0553754   .0099919     5.54   0.000     .0357887    .0749621
       empl3 |   .1772021   .0092229    19.21   0.000     .1591228    .1952813
       empl4 |   .3975509   .0177593    22.39   0.000     .3627381    .4323637
       empl5 |   .3072914   .0183344    16.76   0.000     .2713512    .3432315
      d_2003 |   -.019209   .0074743    -2.57   0.010    -.0338606   -.0045574
       _cons |   .0614897   .0099681     6.17   0.000     .0419498    .0810296
------------------------------------------------------------------------------

. predict yhat (option xb assumed; fitted values) (235 missing
values generated)

. su yhat ,de

                        Fitted values
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -.0071431      -.1471096
 5%     .0111495      -.0884523
10%     .0192287      -.0681077       Obs                8184
25%     .0504399      -.0291871       Sum of Wgt.        8184

50%     .1189307                      Mean           .1511486
                        Largest       Std. Dev.      .1230529
75%     .2225205       .6552632
90%     .3330851       .6632593       Variance        .015142
95%     .4131974       .6661162       Skewness       .9536523
99%     .4793803       .6746188       Kurtosis       3.387368

. log close
       log:  \\Balder\540$\kir\Internett\Annet\ECON 4135\wp2.log
  log type:  text
 closed on:  25 Oct 2007, 15:20:59
------------------------------------------------------------------------------
```