WRITTEN PAPER II (ECON 4135)

October 7, 2009

Individual papers will not be corrected or commented. Students are encouraged to do the work and write a paper, and to discuss it in their colloquium group. The problem set will be discussed in the seminar in week 45.

SkatteFUNN is a system for tax deduction of Research and Development expenses (R&D). It became law in 2003 and applies to any Norwegian enterprise. If a tax deduction is granted to a firm, it gives the right to a 20 percent tax deduction of R&D, limited upwards to 800,000 NOK for one year. If the firm is not in a tax position, i.e., its payable tax is zero, the R&D subsidy is paid out in cash.

To obtain the R&D subsidy through SkatteFUNN, certain formal requirements must be fulfilled: The project must have the character of R&D, and specifically have "a goal to acquire new knowledge, information or experience which is assumed to be useful for the firm with regard to developing new products, services or production processes".

While SkatteFUNN is meant to generate more R&D and change in R&D behavior in enterprises, it may not be effective, either because the firms would undertake the R&D projects anyway, or because the system gives an incentive for creative bookkeeping, whereby ordinary operating expenses are classified as R&D.

We will examine the following problem: What characterizes the firms that obtained R&D subsidies through SkatteFUNN? In this problem set we will look at data for both 2003 and 2004. The 2003 data are contained in the accompanying STATA data file named "manuf2003.dta". The 2004 data are contained in "manuf2004.dta. Our sample consists of (almost all) joint stock companies (i.e., they are not personally owned) in the manufacturing sector. A complete variable list is given in Table 1.

1 . Use STATA to produce summary statistics for the variables in Table 1, except for *orgnr* and *year*. Also present summary statistics for the sub-sample consisting of

Table 1: **Variable list**

Variable name	Description
orgnr	ID (observation unit) number for the firm
RDsubsidy	R&D subsidy from SkatteFUNN, in 1000 NOK
y	Dummy (binary): =1 if RDsubsidy > 0 (and =0 if RDsubsidy=0)
taxposition	Dummy: =1 if the firm has positive payable tax (prior to any tax deduction)
share_high	Share of the firm's workers with at least 17 years of schooling
VA_empl	Value added per employee in 1000 NOK
firmage_10y	Dummy: =1 if the firm's age is 10 years or less
emply	Number of employees
empl1	Dummy: =1 if $emply \leq 5$
empl2	Dummy: =1 if $emply > 5$ and $emply \leq 10$
empl3	Dummy: =1 if $emply > 10$ and $emply \leq 50$
empl4	Dummy: =1 if $emply > 50$ and $emply \leq 100$
empl5	Dummy: =1 if $emply > 100$
year	Observation year: 2003 or 2004

the firms that obtained the R&D subsidy, i.e., with $y = 1$. In particular, give a detailed summary for the variable $RDsubsidy$

2. Our dependent variable will be $RDsubsidy$. Estimate the following regression model with OLS using the 2003 and 2004 data separately.

$$RDsubsidy = \beta_0 + \beta_1 taxposition + \beta_2 share\_high + \beta_3 VA\_empl + \beta_4 firmage\_10y + \beta_5 emply + u$$

Specify the assumptions about the error term u which are needed for consistency and asymptotic normality of the estimator. What is the conditional expectation $E(RDsubsidy|X)$, where $X = (taxposition, share\_high, VA\_empl, firmage\_10y, emply)$? If you use the option *robust* in STATA, which part of your results are changed (and why) ?

3. To try to improve the fit of the models, replace the continuous variable $emply$ with the dummy variables $empl2, \ldots, empl5$. Which of the specifications appears to be the better (the one with $emply$ or the dummy variables)? Interpret the results briefly. What would happen if the dummy variable $empl1$ were also included in the regression? Explain briefly what is meant by multicollinearity. Why is this a particularly relevant problem when the model contains dummy variables? If $emply2$ is excluded instead of $emply1$, how will this affect the results for the parameters of the included variables in the model? (Answer this question without reestimating your models!)

4. Give 99% confidence intervals for the parameters of the variables VA_empl, share_high and firmage_10y based on the 2003 data only. What does it mean that

the degree of confidence is 99%? Which of these variables are significant determinants of the dependent variable *RDsubsidy* at the 1% level of significance? Give a brief interpretation of your findings with regard to these three variables.

5. Let us now turn to the variable *taxposition*, which takes the value one if the firm is in tax position in the observation year, i.e., if it pays tax, and zero if not. Use the results for both 2003 and 2004 and give 99% confidence intervals for β_1 in both cases. Discuss whether β_1 is the same in both years, i.e. whether this parameter is unchanged from 2003 to 2004. Set up an appropriate test statistic and carry out a formal statistical test.

6. One way of using data from both years in one estimation is to apply the *append* command in STATA: Open the data set "manuf2003.data" and run the command "append using manuf2004". Give the new data set – which contains both the 2003 and 2004 data – a new name and save it. Then generate a dummy for the observation year 2003, e.g. by the command "generate d_2003=(year==2003)". Include this time-dummy (d_2003) in addition to the other variables, in a regression using the combined 2003 and 2004 data set. Is the parameter of the time dummy significantly different from zero? Interpret the results! What would happen if a dummy for 2004 was also included in the regression?

7. Let us now look closer at the variable *taxposition*, which takes the value one if the firm is in tax position in the observation year, i.e., if it pays tax, and zero if not. Can you formulate a relevant statistical hypothesis to investigate whether there is any relation between a firm's tax position and the amount of R&D subsidies it gets? Set up a test statistic and calculate the p-value of the test. Use the combined 2003 and 2004 data set described above. What does the p-value say? It appears that there is a very significant, negative relation between tax position and R&D subsidy. Can you suggest an explanation of this finding?

8. In the combined data set for 2003 and 2004, the same firm often appears in both years. May this lead to a violation of any of the basic assumptions behind using OLS? Explain!

9. In view of the summary statistics for the dependent variable *RDsubsidy*, does this variable appear to have a normal distribution? Do you think a linear regression model is appropriate, in view the summary statistics? An alternative model would be a binary regression model with the binary variable y as the dependent variable. However, since we have not yet come to logit and probit models, this will have to wait. Instead, estimate a linear regression model with y as the dependent variable. Do you have to modify any of the main conclusions about the relation between SkatteFUNN and the explanatory variables in view of the new results? Use the *predict* command in STATA

to see whether the predicted values of $E(y|X)$ – where X contains all your regressors – are plausible. What would you say are definitely not plausible values for $E(y|X)$