Question 1

a)

I would use difference in difference to estimate the effect of the debt reduction policy implemented. To do this I need to find the difference between individuals in tiltakssonen and d1993.

$$\left(y_{93}^{Tiltak} - y_{87}^{Tiltaksonen}\right) - \left(y_{93}^{Not} - y_{87}^{Not}\right)$$

Plotting in the effect in the difference in difference above.

$$(890 - 800) - (1130 - 1100)$$
$$= 60$$

The effect of the debt reduction policy implemented in "Tiltakssonen" from 1991 on had an estimated increase of 60 average hours worked per individual.

b)

To formulate an econometric model I will include and interaction term between the dummy Tiltakssonen and d1993. This is because we have two dummies and our sample is a pooled regression.

$$average\ hours = \beta_0 + \delta_0 d1993 + \beta_1 Tiltaksonen + \delta_1 Tiltakssonen * d1993 + u$$

D1993 is a dummy variable for policy while the tiltaksonen is a dummy variable if unit belong to treatment group or not.

$$y_{87}^{Not} = 1100 = \beta_0$$
$$y_{93}^{Not} = 1130 = \beta_0 + \delta_0$$
$$y_{87}^{Tiltak} = 800 = \beta_0 + \beta_1$$
$$y_{93}^{Tiltak} = 890 = \beta_0 + \beta_1 + \delta_0 + \delta_1$$

Calculating the coefficients we get given what is over we get:

$$\delta_0 = 1130 - 1100 = 30$$

$$\beta_1 = -300$$
$$\delta_1 = 890 - 1100 - 30 + 300 = 60$$

$$\widehat{average\ hours} = 1100 + 30 * d1993 - 300Tiltaksonen + 60Tiltaksonen * d1993$$

From the econometric model containing the variables above and the coefficients we can see that the coefficient $\widehat{\delta_1} = 60$ is the same amount as the estimated effect of the debt reduction policy from previous task. Therefore we can test if the policy effect is statistically significant with coefficient δ_1 using a t-test. If we are able to reject the null hypothesis of $\delta_1 = 0$ with a reasonable level of rejection region then the policy effect has an significantly effect on average hours worked per individual. You can also do a one-tail test to see if the policy effect had an positive effect on the dependent variable, $\delta_1 > 0$

To construct a confidence interval, you can use the formula and plotting in the estimated coefficient for δ_1 and its standard error. Note the critical value depends on the sample size and significant level. Example if the sample is 1000 and significance level of 95% meaning a 5% rejection region then critical value is 1.96:

$$-c < \frac{\widehat{\delta_1} - \delta_1}{se(\widehat{\delta_1})} < c$$
$$-c * se(\widehat{\delta_1}) < \widehat{\delta_1} + \delta_1 < se(\widehat{\delta_1}) * c$$
$$\widehat{\delta_1} - c * se(\widehat{\delta_1}) < \delta_1 < \widehat{\delta_1} + se(\widehat{\delta_1}) * c$$

If our hypothesis value is within the confidence interval then we fail to reject the null hypothesis. Outside we reject the null hypothesis given the testing value of coefficient and the significance level.

c)

I would extend my approach by including age and complete in my econometric model. The ministry is worried about omitted variable problem. It is when we estimate a model that is not the true model. We are however estimating a model without variable that might affect the dependent variable and our independent variables. This leads to either a positive or downward bias. Our estimators isn't unbiased.

To account for the suggestions from the Ministry, I would include dummies **Age and Complete** in our econometric model. I would include age as an separate variable while complete as an interaction term with the other dummies. This alteration of our model should take into account a possible systematically difference in terms of age and basic education in "Tiltakssonen".

$$
\begin{aligned}
&average\ hours \\
&\quad = \beta_0 + \delta_0 d1993 + \beta_1 Tiltaksonen + \beta_2 Age + \beta_3 Complete \\
&\quad + \delta_1 Tiltakssonen * d1993 + \delta_2 Tiltasonen * Complete \\
&\quad + \delta_3 d1993Complete + \delta_4 d1993 * Tiltaksonen * Compltete + u
\end{aligned}
$$

A simpler model is an interaction term between Tiltaksonen and Complete. Also include age as independent variable

$$
\begin{aligned}
&average\ hours \\
&\quad = \beta_0 + \delta_0 d1993 + \beta_1 Tiltaksonen + \beta_2 age + \delta_3 Complete \\
&\quad + \delta_1 Tiltakssonen * d1993 + \delta_2 Tiltaksonen * Complete + u
\end{aligned}
$$

May test the hypothesis that average hours worked between people finished upper secondary and not finished have been constant across being in tiltaksonen or not.

d)

Since High is another dummy variable, then we can reformulate model in b) to estimate difference in difference in differences. To do this our model we include high as an separate variable and interaction term with the other two dummies.

$$
\begin{aligned}
&average\ hours \\
&\quad = \beta_0 + \delta_0 d1993 + \beta_1 Tiltaksonen + \beta_2 High + \beta_3 TiltaksonenHigh \\
&\quad + \delta_1 Tiltakssonen * d1993 + \delta_2 d1993 * High + \delta_3 d1993 * Tiltak * High \\
&\quad + u
\end{aligned}
$$

$$y_{87,Not,Not\ H} = \beta_0$$

$$y_{93,Not,Not\ H} = \beta_0 + \delta_0$$

$$y_{87,Tiltak,Not\ high} = \beta_0 + \beta_1$$
$$y_{93,Tiltak,Not\ High} = \beta_0 + \beta_1 + \delta_0 + \delta_1$$
$$y_{87,Not,H} = \beta_0 + \beta_2$$
$$y_{93,Not,H} = \beta_0 + \delta_0 + \beta_2 + \delta_2$$
$$y_{87,Tiltak,H} = \beta_0 + \beta_1 + \beta_2 + \beta_3$$
$$y_{93,Tiltak,H} = \beta_0 + \beta_1 + \delta_0 + \delta_1 + \beta_2 + \beta_3 + \delta_2 + \delta_3$$

Difference in difference in difference is:

$$=\left[\left(y_{93,Tiltak,H} - y_{87,Tiltak,H}\right) - \left(y_{93,Tiltak,Not\ High} - y_{87,Tiltak,Not\ high}\right)\right] - \left[\left(y_{93,Not,H} - y_{87,Not,H}\right) - \left(y_{93,Not,Not\ H} - y_{87,Not,Not\ H}\right)\right]$$
$$= [(\delta_0 + \delta_1 + \delta_2 + \delta_3) - (\delta_0 + \delta_1)] - [(\delta_0 + \delta_2) - (\delta_0)]$$
$$= (\delta_2 + \delta_3) - \delta_2$$
$$= \delta_3$$

By including the dummy variable High as an separate and as an interaction term with the other dummies from model 2, we have now taken into account this information to account for the debt reduction policy in "tiltakssonen" depending if the individual has higher education or not.

I can now t-test the coefficient δ_3 for the policy given the new information from a pooled regression by OLS.

For model b and d, it assumes that the parallel trend assumption is satisfied. This assumes that if the people in tiltakssonen is gone, the estimated effect for people in the treatment group (tiltakssonen) should be the same for what we estimate to be the effect for people outside tiltakssonen.

Question 2

a)

The estimated coefficients in column (1) tell the following:

If husbands education increases by one year, then we expect total income of husband and wife to increase by 4.39%, all else equal in the model. The reason for 4.39% is that the dependent variable is in logarithmic form. For interpreting the coefficients we need to multiply by 100 to estimate the effect on dependent variable.

For a one year increase in wifes education, total husband and wife income is estimated to increase by 3.9% (100%*0.039), all else equal in model.

The estimated coefficients tells me that the more education the husband and wife has, the more increases total income for husband and wife.

b)

The estimated impact of husband´s education differ between the column because we have omitted wifes education in column (2). This is a omitted variable problem. As seen from table 2, wedu has a positive correlation between hedu and lfamince. Excluding wedu in column (2) leads to hedu to overestimate its effect on lfaminc. The interpretation can be seen as:

$$lfaminc = \beta_0 + \beta_1 hedu + \beta_2 wedu + u$$

Column (2) estimate:

$$lfaminc^* = \beta_0^* + \beta_1^* heduc$$

β_1^* should be unbiased, but it is unbiased since we have omitted wedu. We can write the relationship between β_1, β_1^* and β_2 as.

$$\beta_1^* = \beta_1 + \beta_2 \delta_1$$

δ_1 comes from if you regress wedu on hedu. From table 2 we see that the correlation is positive, meaning the sign in front of δ_1 is positive. This leads as mentioned a positive bias, since β_2 is also positive from column (1). That is why the coefficient for hedu in column (2) is higher than in column (1). It is overestimating the effect of another year of husband education on total income.

c)

Interpretation:

For another year of education for the husband, total income for husband and wife is expected to increase by 0.48% (0.0048*100), all else equal in the model.

If edutot increase by another year of education, total income for husband and wife is expected to increase by 3.9% (0.039*100), all else equal in the model.

The constant

d)

Husband and wife´s education has the same impact on family income. Setting up the hypothesis.

$$H_0 = \beta_1 - \beta_2 = 0$$
$$H_A: \beta_1 - \beta_2 \neq 0$$

$$\beta_1 - \beta_2 = \theta$$

Rewrite

$$\beta_1 = \theta + \beta_2$$

Plot this in the estimated model

$$\widehat{lfaminc} = \beta_0 + (\theta + \beta_2)hedu + \beta_2 wedu + u$$
$$\widehat{lfaminc} = \beta_0 + \theta hedu + \beta_2(hedu + wedu) + u$$

We know from previous task that edutot = hedu+wedu. Therefore, the model over is the one estimated in column (3). Using the coefficient in front of hedu in column (3), we can test if husbands and wife´s education has the same impact on family income.

Using a t-test

$$\frac{0.0048 - 0}{0.0182} = 0.26 = T - stat$$

The critical value is 1.96 with a rejection region of 5% and observations of 428. Comparing it with the T-stat of 0.26, the critical value is larger. This means we fail to reject the null hypothesis with a 5% rejection region. With a larger region, we will still fail to reject. It seems we can conclude that husband and wife´s education has the same impact on family income. This is what should be expected in todays world.

e)

If the number of kids in the family younger than 6 years old increase by one child, the estimated effect on total income of husband and wife is a decrease of 17.33% (-0.1733*100), all else equal in the model.

Impact of husband´s education on total income increase from 0.0439 to 0.0448. The explanation for this might be that it is the wife that takes care of the child. For the wife to take care of the child, the wife might drop taking higher education or reduce hours working. Hours worked isn’t taken into account in the model and isn’t captured by the amount of education the wife has.

Impact of wife´s education increases also from 0.0390 to 0.0421. The increase might be a result of wife deciding to raise a child drops taking more years of education. Wife’s that takes other year of education will give a greater impact of total income.

Note. The standard deviation for the hedu and wedu reduces in (4) compared to (1). Also R-squared increases in (4). Therefore, column (4) is a better model and gives a better estimator of hedu and wedu to lfaminc.

f)

I can test whether the model estimated in (1) is a correct specification by doing testing that the coefficients in from of $\widehat{lfaminc}^2 \text{ and } \widehat{lfaminc}^3$. I will call these δ_1 and δ_2.

$$H_0: \delta_1 = \delta_2 = 0$$
$$H_A: \delta_1 = \delta_2 \neq 0$$

If the we fail to reject the null hypothesis then $\delta_1 = \delta_2 = 0$, which means that column (1) is not a misspecification. Now for the test using F-test.

$$F - stat = \frac{R_{ur}^2 - R_r^2}{1 - R_{ur}^2} * \frac{n - k - 1}{q}$$

Unrestricted model is table 4, while restricted is column (1) in table 3. K = 4 since we have four independent variables and q = 2 since $\widehat{lfaminc}^2$ and $\widehat{lfaminc}^3$

$$F - stat = \frac{0.179 - 0.171}{1 - 0.179} * \frac{428 - 4 - 1}{2}$$

$$F - stat = 2.06$$

With a rejection region of 10%, the critical value is 2.329. Since F-stat < C, we fail to reject the null hypothesis with a 10% rejection region. It seems we can conclude that there isn’t a misspecification in column (1).

g)

Our model

$$\ln(w_{it}) = \beta_0 + \beta_1 hedu_{it} + \beta_2 wedu + \alpha_i + u_{it}$$

α_i = **unit specific** component representing unobserved variables which are constant over time

u_{it} = **idiosyncratic** component representing unobserved variables that varies across units and over time

$$(2)\ \Delta \ln(w_{it}) = \beta_1 \Delta hedu_{it} + \beta_1 \Delta wedu_{it} + \Delta u_{it}$$

The problem with this statement is that education usually is the same for each individual. If you were to get data for the same families for more than one period, you will get a panel data set, see model over. With this set, you can do a first difference estimation or do a fixed effect estimation.

For example doing first difference, model (2), you remove the unit specific error term and time-constant variables if they were in the model. (Time-constant can be female dummy.) This is probably what the student thinks is good with this statement, but the problem lies with our variables of interest. When you get married you usually don’t take more education. They are ready to start a family rather than focusing on their career. This might lead to small or no variation in $hedu_{it}$ or $wedu_{it}$ over time. This makes it hard to find the effect of education has on total family income. The same goes if you were to do a fixed effect estimation.

Even if there was enough variation to estimate effect of education on total income, we still need Δu_{it} to be uncorrelated with our variables of interest. If not, then our estimation isn't unbiased and consistent estimator for β_1 and β_2.