NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 1

Emnekode / *Subject:* Søk3001 Antall ark / *Number of pages:* 37

Denne kolonne er forbeholdt sensor

This column is for external examiner

1

a) The coefficients in a ~~regressi~~ linear regression model tell us how much the dependent variable (left-hand ~~side~~ variable) increases due to a one unit increase in the explanatory variable in question (Personal exemption for example), holding all other things constant.

$$FR_t = 55.944 + 0.178\,PE_t + 0.0035\,M\&AI_t - 68.12\,Unemp_t + 0.393\,Inf.Mor_t + 964.13\,Immigration_t + 15.427\,FW - 25.383\,Pill - 29.479\,WWII - 0.843\,T$$

Fertility Rate (FR_t) is expected to increase by 0.178 children born per 1000 women due to a one dollar increase in personal tax exemption (PE_t), all else being equal.

A one dollar increase in personal income per family (M&AI), net of female earnings, is expected to increase the nr. of children born by 0.0035 per 1000 women, all else being equal.

If the share of unemployed people ($Unemp_t$) increases by 0,01 (1% point) the fertility rate is expected to decrease by $0,01 \cdot 68,12 = 0.6812$ children born per 1000 women, ceteris paribus (all else being equal).

Svanemerket trykksak fra Skipnes Kommunikasjon Lisensnr. 241 731

Den gule kopien beholder du / *Please keep the yellow page*

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 2

Emnekode / *Subject:* SØK3007 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

If Infant mortality (Inf. Mor) increases by one child who dies after birth, the fertility rate is expected to increase by 0.393 children born per 1000 women, ceteris paribus.

A 0.01 increase (1 percentage point) in the share of immigration (people who are foreign born) is expected to increase fertility by 9,64, which seems to be very large increase, ceteris paribus.

If female wages increase by 1 dollar, the fertility rate is expected to increase by 15.427 children born per 1000 women. Also seems to be an unlikely large effect if all else is held constant.

After the pill was introduced (1963-1984), the fertility is expected to be 25,383 nr. children per 1000 women lower than before the pill was introduced. ~~Makes sense as birth control~~

During WW II the fertility is expected to be 29,419 lower than before or after WW II, ceteris paribus.

The coefficient of the time trend is -,843, meaning that fertility is expected to drop by 0.843 for each year that passes, ceteris paribus

$R^2 = 0.829$, ~~[illegible]~~ telling us that 82.9% of the variation in fertility rate is explained by our regression model.

This model assumes that personal tax exemptions will have an immediate effect on fertility. ~~The same goes for all variables in the model~~ This is highly unlikely as due to the natural pregnancy, "producing" a child takes at least nine months. In addition, people ~~do not~~ probably spend time on deciding to have a baby, and it could take some time to conceive a child.

It is more likely that the variables have a lagged effect (if any) on the fertility rate.

Is PE_t significant?

$H_0: \beta_1 = 0$ vs. $H_0: \beta_1 \neq 0$

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)} = \frac{0.178}{0.0977} = 1.82 \sim t_{n-k-1} \quad df = 62$$

Reject H_0 if $|t_{\hat{\beta}_1}| > c$, $c_{5\%} \approx 2$ $c_{10\%} \approx 1.67$.

We would reject H_0 at 10% but fail to do so at 5%. The effect is weakly statistically significant.

NTNU

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 4

Emnekode / Subject: Søk 3007 Antall ark / Number of pages: 37

Denne kolonne er forbeholdt sensor

This column is for external examiner

b) The model in column (1) assumes:

TS 1. The model is linear in it's parameters

~~2. Random sampling (~~

TS. 2. No perfect collinearity and enough variation in the explanatory variables

TS. 3. Zero conditional mean → $E(u_t \mid X)$, strict exogeneity. The error term must be uncorrelated with all x's at all times.

Under TS. 1 - TS. 3, the OLS estimators are unbiased and consistent.

TS. 4 - Homoskedasticity - constant variance in the error term

TS. 5 - No serial correlation - $Cov(u_t, u_s \mid X) = 0 \quad \forall t \neq s$

The error terms in two different time periods (s and t) must be zero, given all x's at all time periods.

In this question, it is TS. 5 which is of most interest. Column (2) is way of solving for the issue of serial correlation, where the equation is transformed to ~~allow~~ solve for serially correlated error terms. (Described in c)).

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 5

Emnekode / Subject: SØK 3007

Antall ark / Number of pages: 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

If the error terms are in fact serially correlated, this will violate the 5'th assumption, and cause the standard errors of OLS-estimators to be incorrect.

This will render t-tests, F-tests, Confidence intervals invalid, making us unable to perform inference. Also, the model will no longer be BLUE (best linear unbiased estimator) as the standard errors are incorrect.

The assumptions previously stated are called the strict assumptions, and they are very rarely valid. As $n \to \infty$, a set of weaker assumption will be valid, which are less strict and more likely to hold. (See next page).

The transformation done to correct for serial correlation will likely violate the set of strict assumptions, as TS.3 → $E(u_t \mid X) = 0$. ~~This means that lagged variables~~

Mai 2018

Svanemerket trykksak fra Skipnes Kommunikasjon Lisensnr. 241 731

Den gule kopien beholder du / Please keep the yellow page

Weaker TS assumptions.

TS.1' - Linearity, stationarity and weak dependence

TS.2' - No perfect collinearity & enough variation in X

TS.3' - Contemporaneous exogeneity $E(u_t | \mathbf{X}_t) = 0$

TS.4' - Homoskedasticity $\rightarrow$ $Var(u_t | \mathbf{X}_t) = \sigma^2$

TS.5' - No serial correlation $\rightarrow$ $Cov(u_t, u_s | \mathbf{X}_t, \mathbf{X}_s) = 0$ $\forall s \neq t$

Under TS.1' - TS.3', the OLS estimator is consistent, but only in large samples. This is somewhat a contradiction as in time series data we do not generally have very large samples. In this case we have $n = 72$, which may be large enough to assume the asymptotic assumptions hold, and ~~[illegible]~~ thus allow us to use the central limit theorem and law of large numbers.

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 7

Emnekode / *Subject:* SØK3007 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

Testing for serial correlation:

Assume $u_t = \rho u_{t-1} + e_t$, e_t is iid $\sim N(0,1)$

We want to test $\rho = 0$, but we do not observe the error term u_t directly.

We use the OLS residuals ~~to~~ as estimates,

$$\hat{u}_t = \Sigma(y_t - \hat{y}_t).$$

We then regress the model:

$$\hat{u}_t = \rho \hat{u}_{t-1} + e_t$$, with OLS.

This gives us:

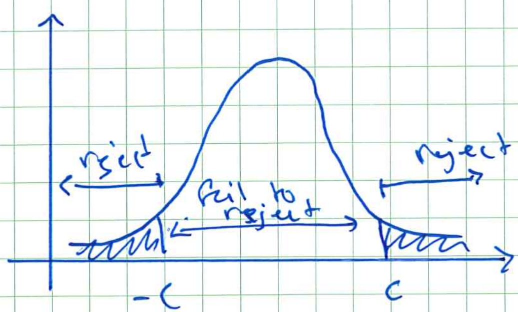
$$\hat{u}_t = \hat{\rho} \hat{u}_{t-1}$$

We can now test for serial correlation:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

$$t_{\hat{\rho}} = \frac{\hat{\rho} - 0}{se(\hat{\rho})} \sim t_{n-k-1}$$



Reject H_0 if $|t| > c$.

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 8

Emnekode / Subject: SØK3001

Antall ark / Number of pages: 31

Denne kolonne er forbeholdt sensor
This column is for external examiner

c)

Assume we have a simple function:

(1) $y_t = \alpha_0 + \beta_1 X_t + u_t$, where u_t is correlated with previous values of u.

We lag the model by one period:

$y_{t-1} = \alpha_0 + \beta_1 X_{t-1} + u_{t-1}$ $| \cdot \rho$, where ρ is ~~the~~ the factor u_t is correlated with u_{t-1} by

(2) $\rho y_{t-1} = \rho \alpha_0 + \rho \beta_1 X_{t-1} + \rho u_{t-1}$

So, $u_t = \rho u_{t-1} + \text{error} \leftarrow \text{iid} \sim N(0,1)$

We subtract (2) from (1):

$$\underbrace{y_t - \rho y_{t-1}}_{\ddot{y}_t} = \underbrace{(1-\rho)\alpha_0}_{\ddot{\alpha}_0} + \beta_1 \underbrace{(X_t - \rho X_{t-1})}_{\ddot{X}_t} + \underbrace{(u_t - \rho u_{t-1})}_{\ddot{u}_t}$$

$$\Rightarrow \ddot{y}_t = \ddot{\alpha}_0 + \beta_1 \ddot{X}_t + \ddot{u}_t$$

The model is called FGLS (feasible generalized least squares) as we do not observe the correlation factor ρ directly, so it must be estimated.

The ρ is estimated by:

1) Running a OLS regression on the original model

$\Rightarrow$ Obtain the residuals from this regression ($\hat{u}_t$)

2) Regress the residuals on the lagged residuals ($\hat{u}_{t-1}$) and other explanatory variables.

$\Rightarrow$ Obtain an estimate of ρ, $\hat{\rho}$.

Denne kolonne er forbeholdt sensor

This column is for external examiner

Such a transformation will not yield unbiased estimators ($\hat{\beta}_j$), but will yield efficient estimators, ie, the estimators with the lowest variance.

If the error terms are not ~~to~~ serially correlated (unlikely), and the strict TS assumptions hold (highly unlikely), OLS on model (1) will yield BLUE estimates, ie the best, linear unbiased estimators.

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} \hat{r}_{ij} y_i}{\sum_{i=1}^{n} \hat{r}_{ij}^2},$$

$\hat{r}_{ij}$ is the residual from regressing x_{ij} on all the other explanatory variables.

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 10

Emnekode / *Subject:* SØK 3001

Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

d) The variable is the average of personal exemption in year t, and it's two predecessing values PE_{t-1} and PE_{t-2}.

To find the long run impact of ~~fertility~~ tax exemptions:

$$\{\beta_4 \cdot X_{4t} \ldots\}$$

$$FR_t = \beta_0 + \beta_1 PE_t + \beta_2 PE_{t-1} + \beta_3 PE_{t-2} + X\beta + u_t$$

Set $PE_t = PE_{t-1} = PE_{t-2} = \overline{PE}$:

$$FR_t = \beta_0 + (\beta_1 + \beta_2 + \beta_3)\overline{PE} + X\beta + u_t$$

$$LRP = \frac{\partial FR_t}{\partial \overline{PE}} = (\beta_1 + \beta_2 + \beta_3)$$

Assuming $\overline{PE}$ is the average value of the PE_t ~~with~~ and it's two lags, we can say the proposed variable does capture the long-run impact of personal tax exemptions on fertility.

Including this variable would allow us to test whether the ~~effect~~ Long run effect is statistically significant or not.

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 11

Emnekode / *Subject:* SØK 3007

Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

$H_0: (\beta_1 + \beta_2 + \beta_3) = 0$ restricted $\rightarrow R_r^2$

$H_1:$ ~~Rot~~ $\cancel{\neq 0}$ unrestricted $\rightarrow R_u^2$

$(\beta_1 + \beta_2 + \beta_3) \neq 0$

~~Reject~~

$\cancel{F = \frac{R_u^2 - R_r^2}{(1 - R_u^2)} \cdot \frac{n-k-1}{q} \sim F_{q,n-k-1}}$

$n = 68$, $k = $ ~~#~~ variables $= 9$, $q = $ ~~# restrictions =~~

$$t = \frac{(\widehat{\beta_1 + \beta_2 + \beta_3}) - 0}{se(\widehat{\beta_1 + \beta_2 + \beta_3})} \sim t_{n-k-1}$$

$$t = \frac{0.191}{0.0477} = 4.00 \sim t_{68-9-1} \approx t_{58} \approx t_{60}$$

Reject if $|t| > c$, $c_{10\%} = 1,671$, $c_{1\%} = 2,66$

$\Rightarrow$ Since $|4.00| > 2,66$, we reject H_0, and can conclude the long run effect of tax exemption is significant at the 1% level.

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 12

Emnekode / *Subject:* SØK 3001

Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

e)

Test whether the time trend is insignificant:

$$FR_t = \beta_0 + \beta_1 T.PE_t + \beta_2 M\&AI_t + \beta_3 Unemp_t + \beta_4 Inf.Mor_t$$
$$+ \beta_5 Immigration_t + \beta_6 FW_t + \beta_7 Pill_t + \beta_8 WWII_t$$
$$+ \beta_9 \cdot TimeTrend + u_t$$

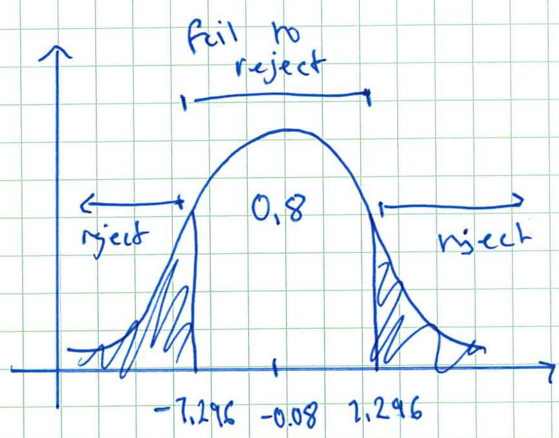
$H_0: \beta_9 = 0 \rightarrow$ insignificant time trend

$H_1: \beta_9 \neq 0$

$$t_{\hat{\beta}_9} = \frac{\hat{\beta}_9 - 0}{se(\hat{\beta}_9)} = \frac{-0.377}{4.77} = -0.08 \sim t_{58} \approx t_{60}$$

$C_{20\%} = 1,296$

$\Rightarrow$ Since $|-0.08| < 1.296$, we fail to reject H_0 even at the 20% significance level. Cannot claim that the time trend is significant



ONTNU

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 13

Emnekode / Subject: SØK3001

Antall ark / Number of pages: 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

Stationarity requires:

1) Constant mean $E(y_t) = E(y_{t+h})$

2) Constant variance $Var(y_t) = Var(y_{t+h}) = \sigma_y^2$

3) The covariance between two observations only depends on the distance between them (h):

$$Cov(y_t, y_{t+h}) = f(h)$$

function of distance between them, not time t.

Since tt is insignificant, y_t does not follow a clear time trend, but it does not necessarily mean that y_t is a stationary process.

If y_t follows a unit root process as the random walk, which violates stationarity,

$$y_t = \rho y_{t-1} + error_t \quad , where \ \rho = 1$$

This gives $E(y_t) = E(y_{t+h})$, 1) is fine, but will give $Var(y_t) \neq Var(y_{t+h}) = f(t)$, and violates the stationarity assumption causing inconsistent estimators:

$$\underset{n \to \infty}{plim}(\hat{\beta}_1) \neq \beta_1$$

Mai 2018

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Den gule kopien beholder du / Please keep the yellow page

First difference:

$$y_t - y_{t-1} = \beta_1 (PE_t - PE_{t-1}) + \beta_2 (MAAI_t - MAAI_{t-1})$$

$$+ \beta_3 (unm_t - unm_{t-1}) + \beta_4 (InfMor_t - InfMor_{t-1})$$

$$+ \dots + (u_t - u_{t-1})$$

$$\Delta y_t = \beta_1 \Delta PE_t + \beta_2 \Delta MAAI_t + \dots + \Delta u_t$$

=> This eliminates the potential violation of stationarity.

Denne kolonne er forbeholdt sensor

This column is for external examiner

f)

Total Chid tax subsidy

In model (5), we have now changed from only personal tax exemption, to the total child tax subsidy, as the explanatory variable of interest.

It may be the case that the variation in total child tax subsidy may very less over time than the personal tax exemption alone. Perhaps ~~than~~ a low personal tax exemption is somewhat offset by a higher earned income tax credit in a period.

~~An~~ The independent variable of interest no longer captures the exact same effect, as we are including more types of tax benefits in (5). This could mean that the correlation between this new variable and the other explanatory variables to be higher than previosly. The estimator in a regression model with multiple explanatory variables will become lower as correlation between explanatory variables increases:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} \hat{r}_{ij} y_i}{\sum_{i=1}^{n} \hat{r}_{ij}^2}$$

, $\hat{r}_{ij}$ is the residual from regressing X_j on all other X's.

Denne kolonne er forbeholdt sensor

This column is for external examiner

g)

As model (1) assumes only a contemporaneous effect of PE on fertility, this seems unlikely and we would not choose this model as the preferred model.

Models (1) to (3) all include time trends which will cause the R^2 to be ~~very~~ very high. However since model (1) ~~and~~ does not have the same dependent variable, comparing R^2 does not makes much sense.

Model (2) corrects for the potential issue of serial correlation by transforming the model with FGLS. Model (3) also does this, but allows us to see the long-run effect of personal tax exemption, in addition to correcting for ~~pot~~ serial correlation in the error term. Thus I would prefer model (3) to ~~model~~ both model (1) & (2).

In these three models ((1),(2)(3)), $\hat{\beta}_1$, the effect of PE on fertility are significant at the 1% level, while model (1) gives significance of $\hat{\beta}_1$ only at the 10% level.

Denne kolonne er forbeholdt sensor
This column is for external examiner

Model (4) and (5) take the first difference of the models, which will correct for the potential violation of stationarity (TS.1'). ~~This removes~~

In model (4) the effect of PE is found to be significant at the 5% level, with

$t_{\hat{\beta}_1} = \frac{(-0.084)}{0.042} = -2$, $c_{5\%}^{df=60} = 2.0$

$\Rightarrow$ The p-value is $\approx 5\%$ $\rightarrow$ reject the null hypothesis of $\beta_1 = 0$.

Model (5) gives an $R^2 = 0.103$ while (4) yields $R^2 = 0.203$. In addition, the effect of total tax subsidy is not found to be significant at the 5% level, since $t_{\hat{\beta}_1} = -1.167$ ~~reject~~ H0 at 20% level.

Throughout all models, the effect of increased tax incentives are expected to be rather small impacts on fertility. ~~as~~ In each model, fertility is expected to increase with less than 0.2 children per 1000 women, meaning around 200 children per million women.

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 18

Emnekode / Subject: SØK 3007 Antall ark / Number of pages: 31

Denne kolonne er forbeholdt sensor
This column is for external examiner

I would prefer to use model (5) among the presented models. This controls for most ~~factors~~ potential issues which violate our assumptions needed to correctly estimate the relationship between tax incentives and fertility, although the model has a low R^2 of 0.103 and does not allow us to conclude that tax incentives do in fact have a significant effect on the populations decisions to have babies.

~~Ad~~ The coefficients on unemployment, infant mortality, pill are not statistically significant. ~~Unless these are correlated with to~~ Dropping these variables may ~~improve the R^2 of the~~ give ~~the~~ lower stand. errors for total child tax subsidity, and possibly give significant effect of this on fertility

Denne kolonne er forbeholdt sensor
This column is for external examiner

2

a) Violent Crime = VC

Police per capita = Pol

State prisons per capita = Pris

Unemployment rate = Unem

State income = Inc

Effective abortion rate = Abor

City population = pop

Percentage black = PBlack

$$(1): \log(VC_{ct}) = \beta_0 + \beta_1 \log(Pol_{ct}) + \beta_2 \log(Pris_{ct}) + \beta_3 Unem_{ct} + \beta_4 Inc_{ct} + \beta_5 Abor_{ct} + \beta_6 pop_{ct} + \beta_7 PBlack + \delta_t T + u_{ct} + a_c$$

unobserved heterogeneity

time dummies

This a pooled OLS model where we have the following assumptions to pin down the casual impact.

Pols 1: Linearity

Pols 2: Random sampling, the sample must be randomly drawn from the population at each time period ~~ct~~

Pols 3: No perfect collinearity and enough variation in **X**.

Pols 4: Zero conditional mean → $E(u_{ct} \mid X_{c,t}) = 0$

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 20

Emnekode / *Subject:* SØK 3001 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

In this case, for OLS to yield unbiased estimates of the regression coefficients we have ~~place~~ fairly strong assumptions. ~~OR 2 the~~

POLS.2 requires that the sample is randomly drawn at each time period. Here we have data on the same 122 cities over time, so the observations across time are likely to be correlated with ~~one~~ ~~and~~ each other. ~~If a to~~ This could be due to some city-specific factor which is not controlled for in a POLS, often called unobserved heterogeneity ~~factor~~. (a_c).

This will violate the POLS.2 → and cause us to incorrectly estimate the effect of police presence on crime in city c at time t.

Further, POLS.4 is also likely to be violated. There are three potential cases in which we have endogeneity, ie $E(u_{ct} | \mathbf{X}_c, a_c) \neq 0$.

1) Measurement error in any explanatory variables
2) Omitted variable bias
3) Simultaneity bias.

POLS.4 requires that the error term u_{ct} must be uncorrelated with all the explanatory variables at all time periods $(\mathbf{X}_c)$.

Denne kolonne er forbeholdt sensor

This column is for external examiner

Measurement error in x: Consider the simple ~~case~~ model.

true model: $y_i = \beta_0 + \beta_1 X_i^* + u_i$

We measure x with an error:

$$X_i = x_i^* + e_{i1} \iff x_i^* = x_i - e_{i1}$$

This gives us:

$$y_i = \beta_0 + \beta_1(X_i - e_{i1}) + u_i$$

$$y_i = \beta_0 + \beta_1 X_i + v_i \quad , \quad v_i = u_i - \beta_1 e_{i1}$$

OLS minimizes the sum of squared residuals:

$$\min \sum_{i=1}^{n} (\hat{u}_i)^2 = \min \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

FOC:

$$(1) \quad \min_{\hat{\beta}_0} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \to \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$(2) \quad \min_{\hat{\beta}_1} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \to \sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$(1): \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad , \quad \bar{y} = \frac{1}{n} \sum y_i \quad , \quad \bar{x} = \frac{1}{n} \sum x_i$$

$$(2): \quad \sum (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) x_i = 0$$

$$\Rightarrow \sum (y_i - \bar{y})(x_i - \bar{x}) = \hat{\beta}_1 \sum (x_i - \bar{x})^2$$

$$(3) \Rightarrow \quad \hat{\beta}_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 24.05.18 Side / *Page* 22

Emnekode / *Subject:* SØK3001 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

Inserting for $(y_i - \bar{y})$ in (3) gives us:

$$\hat{\beta}_1 = \beta_1 + \frac{\Sigma(x_i - \bar{x})u_i}{\Sigma(x_i - \bar{x})^2}$$

With measurement error we have that:

$$\hat{\beta}_1 = \beta_1 + \frac{\Sigma(x_i - \bar{x})(u_i - \beta_1 e_{i1})}{\Sigma(x_i - \bar{x})^2}$$

$$E(\hat{\beta}_1|X) = \beta_1 + \frac{E(\Sigma(x_i - \bar{x})(u_i - \beta_1 e_{i1}))}{\Sigma(x_i - \bar{x})^{\cancel{2}}}$$

$$= \beta_1 + \frac{Cov(x_i, e_1 \cdot -\beta_1)}{Var(x)} = \beta_1 + Cov($$

$$E(\hat{\beta}_1) = \beta_1 \cancel{+} \beta_1 \frac{\sigma_e^2}{\sigma_x^2} \rightarrow \text{biased!}$$

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 23

Emnekode / *Subject:* SØK 7001 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

OVB:

There could be omitted variables causing bias. In this case:

$E(\hat{\beta}_1 | X) = \tilde{\beta}_1 + \hat{\beta}_2 \cdot \tilde{S}_1$, where $\hat{\beta}_2$ is the effect of the omitted variable on y

and $\tilde{S}_1$ is covariance between x_1 and the omitted variable x_2.

For this to cause bias, $\hat{\beta}_2 \neq 0$ and $\tilde{S}_1 \neq 0$, if not $E(\hat{\beta}_1) = \tilde{\beta}_1$.

Simultaneity:

It could be that $\log(VC) \rightarrow \log(Pol)$, not only $\log(Pol) \rightarrow \log(VC)$.

Since a city with hig crime rates is likely to need more police, this could cause a simultaneity bias when trying to estimate model (1) with OLS.

In this case, I would say we are likely to not pin down the causal effect due to violations of POLS.2 and simultaneity which violates POLS.3.

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 24

Emnekode / *Subject:* SØK3001 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

b)

(1): $\log(vc) = \cancel{\log(pol)}\beta_0 + \beta_1 \log(col) + \dots + \cancel{\gamma_c C_c} + u_{tc}$

$\delta_t T + \gamma_c C_c$

(2): $\log(vc) = \beta_0 + \beta_1 \log(col) + \dots + \cancel{\delta_t T} + \cancel{\gamma_c C_c} + u_{tc}$

city-specific dummies

$H_0: \gamma_c C_c = 0$ → restricted

$H_1:$ not H_0 → unrestricted

$$F = \frac{R_u^2 - R_r^2}{1 - R_u^2} \cdot \frac{n-k-1}{q} \sim F_{q,\,n-k-1}$$

$n = 2005$, $k = 7 + 20 + 121 = 148$

year dummies city dummies

$q = 121$

$$F = \frac{0.93 - 0.601}{1 - 0.93} \cdot \frac{2005 - 148 - 1}{121} = \underline{\underline{72{,}09}}$$

$C_{1\%}^{df=1856} = 2.32$ (actually lower but do not have a table ~~with~~ for $q = 121$ same conclusion would apply)

Reject if $F > C$, since $72.09 > 2.32$, we reject Ho at extremely high significance levels, far higher than 1%. The effect of the city dummies is significant.

Mai 2018

Svanemerket trykksak fra Skipnes Kommunikasjon Lisensnr. 241 731

Den gule kopien beholder du / *Please keep the yellow page*

Kandidat nr. / Candidate no. 10053

Dato / Date 29.05.18 Side / Page 25

Emnekode / Subject: SØK3004

Antall ark / Number of pages: 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

Such a model is called a fixed effects model, where we have corrected for the issue explained in a) of the unobserved heterogeneity (a_c). So in that sense, this model could be more likely to identify the causal effect of police on crime.

However, we have still not accounted for the likely issue of simultaneity! To solve this issue we need to set up a simultaneous equation model where police is explained by amongst other things police. Consider:

vector of coefficients

$$(1) \quad \log(VC) = \beta_0 + \beta_1 \log(Pol) + \boldsymbol{\beta} \mathbf{Z}_1 + u_1$$

$$(2) \quad \log(Pol) = \alpha_0 + \alpha_1 \log(VC) + \boldsymbol{\alpha} \mathbf{Z}_2 + u_2$$

vector of exogenous variables

Find the reduced form of log(Pol):

$$\log(Pol) = \alpha_0 + \alpha_1(\beta_0 + \beta_1 \log(Pol) + \boldsymbol{\beta}\mathbf{Z}_1 + u_1) + \boldsymbol{\alpha}\mathbf{Z}_2 + u_2$$

$$\log(Pol)\cdot(1-\alpha_1\beta_1) = (\alpha_0 + \alpha_1\beta_0) + \alpha_1\boldsymbol{\beta}\mathbf{Z}_1 + \boldsymbol{\alpha}\mathbf{Z}_2 + u_2 + \alpha_1 u_1$$

$$\log(Po) = \underbrace{\frac{\alpha_0 + \alpha_1\beta_0}{1-\alpha_1\beta_1}} + \underbrace{\frac{\alpha_1}{1-\alpha_1\beta_1}}\boldsymbol{\beta}\mathbf{Z}_1 + \underbrace{\frac{1}{1-\alpha_1\beta_1}}\boldsymbol{\alpha}\mathbf{Z}_2 + \underbrace{\frac{u_2+\alpha_1 u_1}{1-\alpha_1\beta_1}}$$

$$\log(Po) = \pi_{10} + \pi_{11}\boldsymbol{\beta}\mathbf{Z}_1 + \pi_{12}\boldsymbol{\alpha}\mathbf{Z}_2 + v_1$$

To be able to identify (1), we require that at least one of the variable in Z_2 does not also be part of Z_1, and that this variable has a non-zero α.

We would now use this variable in Z_2 which is not in Z_1 as an instrument for log(Pol), to estimate the original structural equation (1), this is a two-stage least squares (2SLS):

Stage 1: Regress log(Pol) on the IV and all other exogenous variables.

$\Rightarrow$ Obtain the predicted value of $\widehat{\log(Pol)}$.

Stage 2: Regress log(VC) on $\widehat{\log(Pol)}$ and the exogenous variables in the structural equation.

$$\log(VC) = \hat{\beta}_0 + \hat{\beta}_1 \widehat{\log(Pol)} + \dots + \hat{\gamma}_c C_c$$

where $\hat{\beta}_1$ would yield the causal effect of log(Pol) on log(VC).

For this to be true, we have additional requirements for the variable: ~~the ones~~ .

1) Validity $\rightarrow$ $Cov(Z, u) = 0$

2) Relevance $\rightarrow$ $Cov(Z, \log(Pol)) \neq 0$

further discussed in c) $\rightarrow$

Denne kolonne er forbeholdt sensor

This column is for external examiner

c)

Weak instruments means that the covariance between the instrument z and the endogenous variable log(Pol) is not sufficient to ~~be~~ correctly estimate log(Pol). This refers to the relevance requirement presented in b):

2) Relevance ⇒ $Cov(z, \log(Pol)) \neq 0$

In table (2) column (4), we see log(firefighters) appears to be used as an IV, since this is not included in the structural equation of log(VC). For this IV to <u>not be</u> categorized as a weak instrument, we must ~~[illegible]~~ reject the null hypothesis that the effect of log(firsefighters) is zero against the critical value $c = \sqrt{10}$:

~~[illegible]~~

$$\log(Pol) = \pi_{10} + \pi_{11}\log(\text{firefighters}) + \ldots + u$$

$H_0: \pi_{11} = 0$ $\quad t = \frac{\hat{\pi}_{10} - 0}{se(\hat{\pi}_{10})} = \frac{0.206}{0.050} = \underline{4,12}$

$H_1: \pi_{11} \neq 0$

⇒ Since $4.12 > \sqrt{10}$, we reject H_0, and can ~~[illegible]~~ conclude log(firefighters)

NTNU

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 28

Emnekode / *Subject:* SØK3007

Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

is a relevant IV, and ~~does not~~ not a weak instrument.

However, we cannot test whether log(firefighters) is valid, ie Cov(log(firefighters), u) = 0. This must be assumed to hold.

~~(Note that if log(Pol) is simultaneously explained as log(VCT), we should have log(~~

Den gule kopien beholder du / *Please keep the yellow page*

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 29

Emnekode / *Subject:* SØK3001 Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor
This column is for external examiner

3

$$y = f(x_1, \ldots, x_k; \beta_0, \beta_1, \ldots, \beta_k) + u$$

$E(u \mid x_1, \ldots, x_k) = 0 \rightarrow$ exogeneity

$V(u \mid x_1, \ldots, x_k) = \sigma^2 \rightarrow$ homoskedasticity

a) To estimate the vector of parameters we could use the maximum likelihood estimation method, which is used to estimate logit and probit models.

~~Here, the estimator is derived as the~~

Here, the parameters are chosen ~~by~~ as the coefficients which maximize the likelihood of observing the values which we do.

In the probit model for example, the function $f(x_1, \ldots, x_k; \beta_0, \ldots, \beta_k)$ is the cumulative distribution function, where the error term u is assumed ~~the~~ to be normally distributed with zero mean ~~and~~ $\sim N(0,1)$.

The logit model uses a logistic model in estimating the probability of y being equal to 1:

$$Pr(y=1 \mid x) = \cancel{F}(x_1, \ldots, x_k; \beta_1, \ldots, \beta_k) = \frac{e^{(x_1, \ldots, \beta_0, \beta_1, \ldots, \beta_k)}}{1 - e^{(x_1, \ldots, \beta_k)}}$$

Mai 2017

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Den gule kopien beholder du / *Please keep the yellow page*

Denne kolonne er forbeholdt sensor
This column is for external examiner

Probit : $Pr(y=1|x) = \Phi(\beta_0 + \beta X)$

Logit : $\Lambda(\beta_0 + \beta X)$.

These models are restricting the output values to lie in the interval [0,1], and these outputs are interpreted as the probability of $y=1$ for a given set of values of the x's.

b) We would like ~~these~~ estimator to allow us to calculate average marginal effects of the variables in the model $\{X_1, \dots, X_k\}$. These marginal effects can then be interpreted the same way as a coefficient is interpreted in a linear probability model, i.e a one unit increase in X_1 is expected to increase the probability of $y=1$ by AME_1. Where AME_1 is the average marginal effect of X_1 on y.

Kandidat nr. / *Candidate no.* 10053

Dato / *Date* 29.05.18 Side / *Page* 31

Emnekode / *Subject:* SØK3007

Antall ark / *Number of pages:* 31

Denne kolonne er forbeholdt sensor

This column is for external examiner

c)

We could empirically check if these average marginal effects are significant with the likelihood-ratio test (LR-test).

~~Here, we would divide the extended AM~~

$$LR_{stat} = n \cdot (\mathcal{L}_r - \mathcal{L}_u) \sim \chi^2_1$$

As shown above, this test has a chi-squared distribution, and the nullhypothesis we wish to test will be rejected if $LR_{stat} > C$, where C is found in the table for chi-squared distributions.