NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 1

Emnekode / *Subject:* SOK 3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

Question 1)

a)

The authors want to test whether human capital facilitates implementation of new technology. They specifically look at the role of education in the industrial revolution in Prussia. They look at two phases in the industrial revolution. The first phase is from 1835 - 1850. The hypothesis is that education played a minor role in this phase as it was mostly about use of existing technology. The second phase is from 1850. In this phase new technology was the driving force of industrialization, and it is thus believed that education will play a larger role in this phase.

The most important econometric challenge is simultaneity. Industrialization level is expected

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

NTNU

Kandidat nr. / Candidate no. 10016

Dato / Date 30.11.19 Side / Page 2

Antall ark / Number of pages: 40

Emnekode / Subject: SØK3514

Denne kolonne er forbeholdt sensor

This column is for external examiner

1)
a)
to affect education level. It is however uncertain in which direction. Increased industrialization can increase demand for unskilled labour and thereby demand for child workers. This will then work in the direction of decreased education level. ~~Industrialization will lead~~ The estimated effect of education of industrialization will then be downward biased. Industrialization can on the other hand lead to increased wealth which is expected to increase the demand for education. The estimated effect of education on industrialization will then be upward biased

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Den gule kopien beholder du / Please keep the yellow page

NTNU

Kandidat nr. / Candidate no.

Dato / Date 30.11.19 Side / Page 3

Antall ark / Number of pages: 40

Emnekode / Subject: SDK3514

Denne kolonne er forbeholdt sensor

This column is for external examiner

1)
a)

Omitted variables can also be a problem as it is challenging to control for all variables correlated with both education level and industrializati

Relying on historical data can often be a challenge in econometric studies, as it is often constructed years later. An advantage of looking at Prussia is that there exists high quality data from different Prussian regions for the variables that we are interested in, from the 19th century.

The authors use an IV to handle the problem of simultaneity. They use education level in different regions in 1816 as an IV for education level in different regions in 1849 and in 1871.

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Den gule kopien beholder du / Please keep the yellow page

NTNU

Kandidat nr. / Candidate no.

Dato / Date 30.11.19 Side / Page 4

Antall ark / Number of pages: 40

Emnekode / Subject: SOK3514

Denne kolonne er forbeholdt sensor

This column is for external examiner

Question 1)

b)

column(1): shows OLS estimates of the effect on education in 1849 on industrialisation in 1849. They also control for other variables in 1849 expected to affect education level and industrialization. The estimated effect is positive and statistically significant. These results are expected to be biased ~~and are provided~~ because of simultaneity, and the motivation for providing them is to compare the IV-results with these OLS-results.

Column(2): shows the first stage regression for the IV later used. Control variables from 1849 is also included. We observe a very high F-statistic at more than 6000. This indicates th

November 2019

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

1)

b)

education level among Prussian regions were highly persistent from 1816 to 1849.

Column (3): shows estimations with the IV from the first stage regressi in column (2). Controls from 1849 are also added. The estimated effec are a bit lower than those provide with OLS, but not by much. With a t-stat of 1,7 the results are either barely statistical signifi (at 5% leve or just close under the threshold of being statistical significant. It is normal for IV-estimates to have higher standard errors than OLS-estim A concern with the estimation in (3) is the high F-stat from column (2), when the instrument variable and the IV are so closely correlated, the IV might

Denne kolonne er forbeholdt sensor

This column is for external examiner

1)
b)
not solve the problem of endogeneity caused by simultaneity. If they are perfectly correlated the IV will not contribute with any variation, and IV and OLS will yield the same results.

Another problem with the estimations in (2) and (3) concerns the validity of the IV. Although ruling out the issue of simultaneity we might have factors in 1816 which are correlated both with education level and the propensity to industrialize in later periods. It is there important to control for such variables to ensure the validity of the IV.

This brings me to column (4) and (5) where pre-industrial development factors such as

NTNU

Kandidat nr. / Candidate no. 100

Dato / Date 30.11.19 Side / Page 7

Antall ark / Number of pages: 40

Emnekode / Subject: SØK3514

Denne kolonne er forbeholdt sensor

This column is for external examiner

1)
b)
Sheep per capita is included both in the first stage regression (4) (called control variables 1816) and in the IV-regression (5). We see from (4) that the IV is still highly significant. The estimated effect of education on industrialization is higher than the effect estimated in (3) and close to the OLS-estimate in (1). The effect is statistically significant at 5% level.

OK

Bra!

November 2019

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

Question 1)

c)

In table two we lookat the second part of the industrial revolution. The same IV is used to instrument education level in 1849, education level in 1816. Control variables from 1816 is added in both first stage and structural equation. See that they have also controlled for industrialization level in 1849 by including share of factory workers in total population 1849. This is done to test whether education level in second part had an individual effect on industrialization, or whether education only had an effect through increased industrial in the first period. We see that the IV is highly relevant with an F-stat of 65. From (7) we see that education level in second part has an additional

Yes!

Denne kolonne er forbeholdt sensor

This column is for external examiner

1

d)

positive effect on industrialization.
A 1 percentage point increase in the literacy rate in 1871 is expected to increase share of manufacturing workers in total population in 1882 by around 0,1 percentage points. ~~[illegible]~~ The effect is statistically significant at 5% level. The effect might seem small but if the share of workers in manufacturing initially is small a 0,1 percentage increase can be substantial. We see that industrialization level in 1849 has a large and statistically significant effect on industrialization in the later period.

Bra!

Denne kolonne er forbeholdt sensor

This column is for external examiner

Question 2)

a)

RDD is a way to measure the causal affect of treatment on an outcome variable.

The study objects will receive treatment when the assignment variable reach a threshold c:

treatment if $X \geq c$

not treatment if $X \leq c$

we will there have discontinuity that provide exogenous variation around the threshold or cutoff point.

The idea is that individuals initially above and below this point will be similar study objects, except for the fact that some of them recieve treatment and some not. It can be viewed as rather random whether an individual is just below or

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

a)

just above, making these objects comparable groups. This will in great extent remove problems caused by omitted relevant variables, since we look at individuals believed to share many characteristics.

~~In an RDD design it is important to also include the assignment variable X, as it is likely that this will have an effect~~

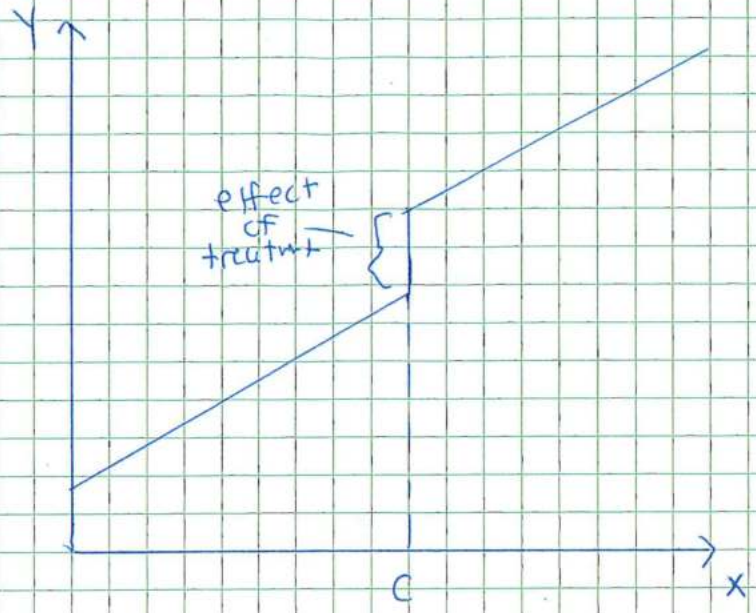
Comparing the difference in mean for the outcome variable for the two groups is therefore believed to give the effect of treatment

This is shown graphically on the next page. The outcome Y makes a jump at $X = c$. This jump is then the effect of treatment

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
a)



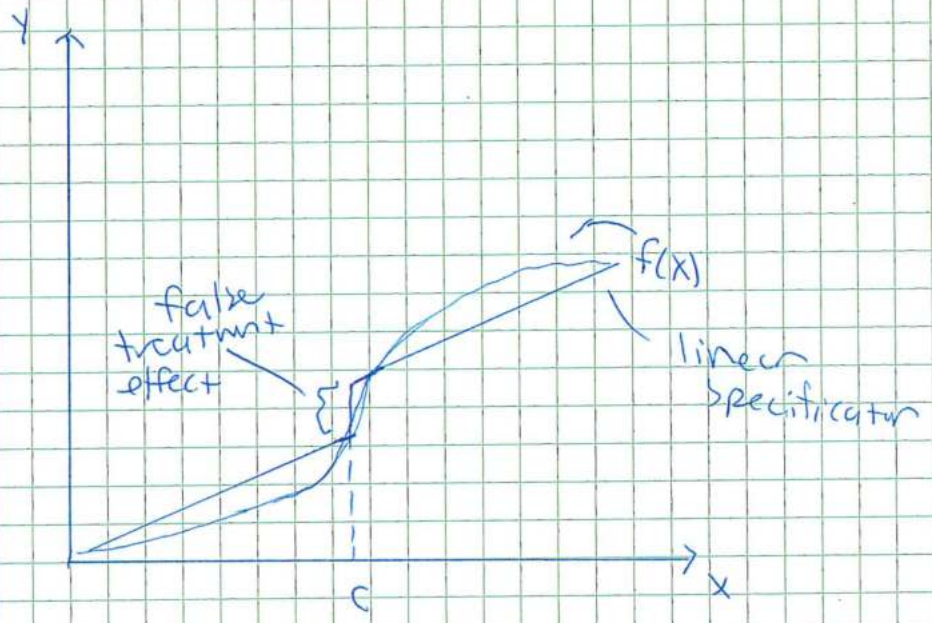
In an RDD it is important to also include the assignment variable X in the estimated regression, as it is likely that this variable will have an independent effect on Y.

It is also important that this function, f(X) is correctly specified. If X also makes a jump at c, we can end up with falsely attributing this to the treatment

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
c)
When it is in fact X that causes this jump



The graph above shows such a situation. f(x) is best specified in polynomials as it makes a jump at c. If f(x) is treated as a linear function, we will estimate a false treatment effect.

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
a)

If individuals can perfectly manipulate whether they receive treatment or not this would invalidate the RDD setup, as treatment then would not be exogenous.

For RDD to be valid other factors affecting y need to develop continuously with x. If these control variables make a jump at the cutoff, this indicates that other factors than the treatment drives the results.

When performing an RDD one can look at both the discontinuity sample or the whole sample. Discontinuity sample means looking at the sample close to the cutoff(s), for example those in the interval ±2 percentage point from cutoff. This approach will lead to less biased results, as the groups

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

a)

just above and just below thresholds are comparable groups. However, the discontinuity sample often includes few observations, making estimates imprecise. Using the whole sample leads to more precise estimates but can cause bias due to reliance on individuals far from cutoff. It is ~~[illegible]~~ therefore a ~~[illegible]~~ trade off between ~~[illegible]~~ precision and unbiasedness.

When using the whole sample a correctly specified control function will help reduce bias. This is not as important when looking at discontinuity samples as linearity is usually a good approximation when looking at small intervalls.

Denne kolonne er forbeholdt sensor

This column is for external examiner

2a)

If we have panel data, fixed effects should be included, ~~when~~ especially when looking at the whole sample, as this will help reduce bias.

An RD design can include several cut off points. ~~[illegible]~~ ~~[illegible]~~ ~~[illegible]~~ will come more back to this in 2b and 2c.

We distinguish between "sharp" design and "fuzzy" design.

Sharp designs mean that treatment is a deterministic function of the assignment variable X. If X reaches the threshold there is a 100% chance of treatment.

A fuzzy design mean that treatment is a probabilistic function of the assignment

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 17

Emnekode / *Subject:* SØK3514

Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
a)

Variable X. Reaching the threshold c increases the probability of receiving treatment. Since treatment is not absolute we therefore need an IV for the treatment variable. This IV need to be relevant and not have an independent effect on the outcome variable.

Bra

November 2019

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

(Question 2)

b)

In Finland there are rules deciding size of local council depending of the population in the municipality. The size of the local council is a deterministic function of population size. We therefore have a basis for a sharp RDD. In Sweden there are minimum requirements for number of council members based on the voter population. We therefore have a basis for a fuzzy RDD. In Finland we can use # of council members directly in the estimation as this is exogenous. In Sweden we uses dummies for the minimum requirements when constructing an IV for actual council size. In both studies we have several cutoffs providing discontinuity

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 19

Emnekode / *Subject:* SOK3514

Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2b)

Pettersen-Lidbom uses panel data for Swedish and Finnish municipalities for several years, providing him a large dataset. Municipalities in Finland and Sweden are rather homogenous, making them attractive study objects compared to conducting similar analysis for example between countries, which usually have different institutions making comparison difficult.

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

b)

Start by looking at Finland. Pettersson-Lidbom provides OLS estimates, which are expected to be biased due to simultaneity. We expect that larger size of government in terms of spending can make it necessary for more council members to be able to handle several and more complex tasks. Another issue is omitted variables we expect that some variables will affect both size of local governments and government spending. OLS gives positive estimates of size on legislature on government spending.

He provides several specifications of the RD-design. He includes several polynomials for the assignment variable f(population size), I think

NTNU

Kandidat nr. / Candidate no. 10016

Dato / Date 30.11.19 Side / Page 21

Emnekode / Subject: SAK3514

Antall ark / Number of pages: 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
b)

up until the fifth polynomial. Going from third to fifth barely changed the estimated results which is a positive sign. He includes municipality fixed effects, which is especially important when looking at the whole sample. When including FE it is essential to have enough time variation in the explanatory variable to be able to obtain precise estimates. In this case it means that a number of Finnish municipalities must have had law induced changes of council size in the period we look at.

RDD with FE estimates are negative and statistically significant when looking at the whole sample.

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)
b)

He also provides estimates for the discontinuity sample: ±25 %. points, ±10, ±5 and ±2,5. All but the last (±2,5 %. points) yield statistically significant results. All results are negative.

For Sweden the IV used has a F-stat of around 10 meaning that it is not very strong but not weak either. This indicates that Swedish municipalities in general does not follow the minimum requirements strictly, but usually have more council members than required. FE's are included as well as specifications of the assignment variable. He only provides estimates for the whole sample, as the discontinuity sample

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 23

Emnekode / *Subject:* SØK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

b)

is very limited. Results are negative but not statistically significant. This can be due to the fact that there are few changes of # of council among Swedish municipalities, providing little within groups variation which is necessary when using FE, and/or that the IV is not that strong.

Pettersson-Lidbom argues that the results from Sweden can be valuable even if not statistically significant, since the estimated results are so similar to those from Finland which are statistically significant

OK

Virker godt forstått

November 2019

Denne kolonne er forbeholdt sensor

This column is for external examiner

Question 2)

c)

I will look at Angrist and Lavy (A
which is an example of fuzzy design and Pettersson Lidbom (2008) (PL
which is an example of sharp design.

AL want to estimate the effect of class size on student performance. This is problematic due to simultaneity (student performance can affect class size) and omitted variables. They make use of an Israeli school norm which provides exogenous discontinuity in class size. The norm says that classes should not have more than ~~[illegible]~~ 40 students. If enrollment is 41 this norms mean dividing it into two classes of 20.5 students. There is not 100% compliance of this norm.

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 25

Emnekode / *Subject:* SOK3584 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2
c)

but it is strong enough so that AL can use it as an IV for predicted class size in an RDD-set up. Exogenous variation around the cutoff points, 40, 80, 120 etc therefore make the basis of the RDD. In this way the problems of endogeneity due to simultaneity and omitted variables are solved. AL discuss whether parents can manipulate whether their children will recieve treatment (attend small classes), but conclude that it is difficult to know whether enrollment will be for example 38 vs 41. This is an example of a fuzzy design with several cutoff points. AL provide both estimates with discontinuity sample and

V

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

c)

with the whole sample. In general they find a positive and statistically significant effect of class size on student performance for most subjects and ~~class levels~~ ~~(1, 2 and 3 grade)~~ grades.

~~Robustness~~

Different specifications of the assignment variable f (class size) is also provided, which barely change the estimated results.

PL provides a rather different RDD set up than AL as he uses sharp design with only one threshold when trying to estimate the causal effect of left party majority on local government spending. Again he uses ~~Swedish~~ ~~municipal~~ panel data for Swedish municipalities

Denne kolonne er forbeholdt sensor

This column is for external examiner

2)

c)

Econometric challenges is there omitted variables. In particular voter preference which will affect both left vote share in local council and government spending as we assume that people voting for left parties want to see increased government spending. PL wants to control for this effect to find the unique effect of left share majority in the local council. Although a multiparty system Sweden have, at least until the 90's experienced rather stable left vs right blocks in local politics. This makes it possible to distinguish between left vs. right majority for most Swedish municipalites for the time period study. Left party majority is then the cutoff which decides treatment.

Denne kolonne er forbeholdt sensor

This column is for external examiner

b)

c)

Left vote share is the assignment variable. The idea behind the RDD is that municipalities that have 48% left share vs 52% left vote share are comparable objects as we, with the RDD setup has been able to control for voter preference in a way that would otherwise have been difficult.

Again PL uses fixed municipality effects as he because of small discontinuity design needs to rely on the whole sample to be able to get statistically significant results. He find that left majority is associated with increased government spending, higher taxes, and lower unemployment,

NTNU

Kandidat nr. / *Candidate no.* 16016

Dato / *Date* 30.11.19 Side / *Page* 29

Emnekode / *Subject:* SØK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

2
c)

Similarities between the two applications provided is that they use RDD to remove problems of endogeneity caused by simultaneity and omitted variables.

They use different RDD setups, fuzzy vs sharp, and one cutoff vs several, relying mostly on whole sample vs. discontinuity sample together with whole sample

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 30

Emnekode / *Subject:* SOK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

Question 3)

a)

According to economic theory increased labour immigration will lead to an increased supply of labor which isolated will lead to reduced wages for natives. Testing this theory is challenging as labor immigration is not an exogenous variable. Simultaneity is an issue as wage level will be a determining factor for the amount of labor immigration. We rarely see that wages in certain sectors decreases as a result of immigration.

Immigrants tend to come in economic booms, and when they expect to meet a good labour market with low unemployment. This works in the direction of increasing wages. If this is not controlled for we can end up underestimating the effect of immigration on native wages

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 31

Emnekode / *Subject:* SØK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

3
a)

To be able to estimate this effect we need information about the contrafactual outcome what would the wage development have been in absence of increased immigration.

Another issue that could invalidate the results if not controlled for is non-random attrition. Some natives will leave their jobs and maybe the entire sector when we have a high influx of immigrants. If they share certain characteristics, for example that they are unskilled workers with low wages, not accounting for them will lead to results that underestimate the effect of immigration on wages. If influx of immigrants also creates more high paying jobs for natives, this will lead to the same issue as the one just mentioned.

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

a)

Brutsberg and Raum look at the effect of labor immigrants in the Norwegian construction sector. They take advantage of the fact that Norway experienced a large influx of Eastern European labor immigrants in 2004 after several Eastern European countries, among them Poland officially joined the EU. Looking at this event will likely reduce some of the problems of endogeneity mentioned previously, as it was not changes within the construction sector in Norway that lead to more people coming in 2004 vs 2003 or 2002. However there is still a reason that they chose Norway over other countries. There was high activity in the Norwegian construction sector at that time and a high demand

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 33

Emnekode / *Subject:* SØK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

a) for more labor.

Therefore to provide exogenous variation in immigrant share, the researchers make use of certification requirement in some sectors, mainly for electricians and plumbers. These requirements made the immigrant shares in these sectors much lower. Immigrants mainly started to work in sectors such as carpenting or painting with few or little restrictions. The idea is that plumbers and electricians will serve as the contrafactual outcome, representing what the wage development would have been without labor immigration. This is then based on the parallel trend assumption, we assume that they would have developed equally in absence of immigration.

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

NTNU

Kandidat nr. / *Candidate no.* 16016

Dato / *Date* 30.11.19 Side / *Page* 34

Emnekode / *Subject:* SOK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

4) They create an IV for immigrant share in certain sectors based on the certification requirements mentioned and the influx of immigrants in the different sectors. The IV is not very strong (relevance) but they use it for some specifications

OK

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

3)

b)

If we compare (1) with (2) we see that in (2) they take advantage of the panel data set, by including individual fixed effects. The estimated effect in (2) is lower than in (1). This can indicate that the problem of non-random attrition is very much present in (1) but at least to some degree dealt with in (2). When including individual FE we use within group variation to estimate the effect. We therefore compare an individual to itself. When using individual FE we thereby control for the problem of low skilled leavers with low pay driving up the average wage when they leave their jobs.

V

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

b)

In (3) and (4) we look at a balanced panel, meaning we only look at individuals who stayed in the sample the whole study period. This is a different approach to tackle the problem of non-random attrition. Estimated effects in (3) and (4) are quite similar to that in (2). We see that estimates barely change when including individual FE in (4). This is a good sign indicating that the balanced panel approach has mostly dealt with the problems of non-random attrition that FE can fix as well.

In (5) and (6) the researchers try to see whether entrants or leavers drive the biased results in (1) we see that dropping entrants

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)
b)

in (6) yield very similar results as (2) and (3), indicating that leavers share certain characteristics which bias the estimated results towards zero. One can assume that lower paid workers are more inclined to leave the sector when there is an influx of immigrants.

It seems as entrants is not the "problem" as estimation in (5) yields results very different from those in (3) and (4).

I would say that (3), (4) and (6) give most credible estimates. They all control for important problem already discussed. It is also a good sign they give such similar estimates. However, none of them control for the fact that increased labor immigration

NTNU

Kandidat nr. / *Candidate no.* 10016

Dato / *Date* 30.11.19 Side / *Page* 38

Emnekode / *Subject:* SAK3514

Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

b)

Can help create more higher paying jobs for natives. One would think that natives are more likely to be promoted to manager jobs. Since this is not controlled for in any of the specifications, one can expect that the results are still biased towards zero.

November 2019

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

3)

a)

The estimated equation would look something like this

$$\text{native wages} = \beta_1 \text{leavers} + \beta_2 \text{entrants} + \alpha_1 \text{immigration share} + \alpha_2 \text{im.share} \cdot \text{leavers} + \alpha_3 \text{in.share} \cdot \text{entrants} + \epsilon$$

statistically significant estimates of α_2 and α_3 would then indicate that wages are expected to develop differently for leavers and entrants vs those who stay when immigration share increases. Including the dummies for leavers and entrants makes it possible to see if they are expected to have lower or higher wages compared to those who stay in absence of immigration.

NTNU

Kandidat nr. / *Candidate no.* 1016

Dato / *Date* 30.11.19 Side / *Page* 40

Emnekode / *Subject:* SØK3514 Antall ark / *Number of pages:* 40

Denne kolonne er forbeholdt sensor

This column is for external examiner

3)

c)

The estimate of α_1 would give us predicted wage change for those staying in the sample the whole time when immigration increases. This would then corresponds to the results for the balanced panel in (3).
If we addition include individual FE we would get the same result as in (4).

Veldig god forståelse

Svanemerket trykksak fra Skipnes Kommunikasjon
Lisensnr. 241 731

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