#1

(a) The coefficients will interpreted in a ceteris paribus manner.

A migrant who has been in the US for 0-5 years is expected to earn 1.255.73 dollars less than a US native, measured in 2010 dollars. Similarily, a migrant who has been in the US for 6-10 years is expected to earn 734.51 dollars less than a US native, measured in 2010 dollars. Lastly, a migrant who has been in the US for 11-20 years is expected to earn 352.93 dollars less than a US native, measured in 2010 dollars.

5

Just by looking at the results in ~~comn~~ column (1), it seems like that the common view that migrants wages converge toward native wages as they spend longer time in the US holds true. Also: They earn less than natives in their first 30 years in the US.

(b) Consider the following relationship $E(\tilde{\beta}_1) = \beta_1 + \beta_6 \delta_1$, where $\tilde{\beta}_1$ represents $\hat{\beta}_1$ in column (1), β_6 is the population coefficient of the after 1890 variable, and δ_1 is the 'true' regression coefficient from the model $Y0\text{-}5_{it} = \delta_0 + \delta_1 \text{after1890} + V_{it}$

If the term $\beta_6 \delta_1$ is nonzero, then our model from column (1) will be biased as $E(\tilde{\beta}_1) - \beta_1 = \beta_6 \delta_1 \neq \beta_1$. Even asymptotically this will hold true as $\text{plim}(\tilde{\beta}_1) = \beta_1 + \frac{\text{cov}(Y0\text{-}5_{it}, u_{it})}{\text{Var}(Y0\text{-}5_{it})}$

and the covariance term will be nonzero as $u_{it} = \delta_1 \text{after1890} + V_{it}$, and $\delta_1 \neq 0$.

If this is true, then our β_1 from column (1) will be wrongly calculated because of omitted variable bias. Since $\tilde{\beta}_1 < \hat{\beta}_1 \leftarrow$ from column (2) I expect the bias to be downward by (lower than the 'true value') the regression results.

Then it is likely that our model in column (1) is misspecified, and the results are invalid as zero conditional mean is broken.

Kandidat nr. / Candidate no. 10046

Dato / Date 20/5/19 Side / Page 3

Emnekode / Subject: SØK 3001

Antall ark / Number of pages: 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

10

(c) In principle it is a good proposal as a female dummy would capture the time-invariant effect of being a female, from a_i. It is important to take note that it must be significant, or else there is no point in including it. But to include it in an FE model would be unnecessary as it eliminates all time-invariant variables. To see this, consider a general model $y_{it} = \beta_0 + \vec{\beta}\,\vec{x}_{it}^T + \gamma_0 \, female_i + a_i + u_{it}$.

Where $\vec{x}$ is a vector/matrix of explanatory variables and $\vec{\beta}$ a vector of β POLS estimators.

The fixed effects estimator, transforms our model with the transformation (time demeaned data)

$y_{it} - \bar{y}_i = \beta_0 - \bar{\beta}_0 + \vec{\beta}(\vec{x_{it}} - \bar{x}_i)^T + \gamma_0(female - \overline{female_i})$

$+ a_i - \bar{a}_i + u_{it} - \bar{u}_i$, from the FE assumptions we have $E(u_i) = 0$. Now since female and a_i are time-invariant, its mean over time is the variable itself. So $female_i - \overline{female_i} = 0$, and $a_i - \bar{a}_i = 0$ and thus

$\ddot{y}_{it} = \vec{\beta}\,\ddot{\vec{x}}^T$, so no I would not include it in column (3).

Denne kolonne er forbeholdt sensor

This column is for external examiner

d) The authors of the paper are claiming that the common view that wage convergence is true is driven by bias caused by measurement error. They are saying that migrant skills are actually lower then what is commonly thought, and that migrants who have experienced negative 'returns' have left the US, and are therefore not included in the data. If (!) this is true, then the authors have a legitimate argument. To see this, consider a ~~struc~~ cross-sectional model

$y_i = \beta_0 + \beta_1 x_i^* + u_i$, and let x be an incorrectly measured variable such that $e = x - x^*$, x^* - true value.

Then, if $cov(x^*, e) = 0$, we can see that

$cov(x, e) = E(xe) - E(x)E(e) \leftarrow$ assumed 0.

$\Rightarrow E((e + x^*)e) = \sigma_e^2$. The problem is that we will have attenuation bias. ~~Then~~ Let $y_i = \beta_0 + \beta_1(x - e) + u_i$

$\Rightarrow y_i = \beta_0 + \beta_1 x_i + \underbrace{u_i - \beta_1 e}_{v_i}$. But now

$cov(x_i, v_i) = E(x_i v_i) - \overset{=0}{E(v_i)} E(x_i) = E(x_i u_i + x\beta_1 e)$

$= -\beta_1 \sigma_e^2$. This is a violation of the zero-conditional mean assumption, so our OLS estimator is biased and therefore not BLUE by Gauss-Markov.

10

Even asymptotically, we see that

$\text{plim}(\hat{\beta}_1) = \beta_1 + \frac{-\beta_1 \sigma_e^2}{\sigma_e^2 + \sigma_{x^*}^2}$ where $Var(X) = \sigma_e^2 + \sigma_{x^*}^2$.

Then we find that

$$\text{plim}(\hat{\beta}_1) = \frac{\beta_1(\sigma_e^2 + \sigma_{x^*}^2) - \beta_1 \sigma_e^2}{\sigma_{x^*}^2 + \sigma_e^2} = \beta_1 \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2}$$

which is biased toward zero as $\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e} < 1$.

Now, I find this claim that $cov(X,e) \neq 0$ is reasonable as it makes sense that migrants who perform badly wage-wise leave the country. The question is then if this actually is true.

So, part 2 of the question:

It depends on what we define as cohort quality. We can (think) argue that wildly different cultures, assuming this does not fall under quality, can ~~be~~ significantly explain wage differences because of discrimination or ability to ~~integrate~~ assimilate to the US culture. Other things, such as ~~names~~ ~~relat~~ name frequency relative to the US population can also explain differences in earnings. Country of birth and age can be relevant too.

#2

(1) $inf = \beta_{10} + \alpha_1 open + \beta_{11} lpcinc + \beta_{12} oil + u_1$

(2) $open = \beta_{20} + \alpha_2 inf + \beta_{21} lpcinc + \beta_{22} lland + u_2$

(a) From do-file

We regress $inf = \gamma_0 + \gamma_1 oil + \gamma_2 lpcinc + \gamma_3 lland + \varepsilon$

and collect $\hat{\varepsilon}^2$. Then we estimate the model

$$\hat{\varepsilon}^2 = \lambda_0 + \lambda_1 oil + \lambda_2 lpcinc + \lambda_3 lland$$

We perform the Breusch-Pagan test with

hypotheses $H_0: \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \vec{0}$ against H_1: not H_0 at $\alpha = 5\%$.

~~At $\alpha = 5\%$, the critical value is~~

The $F\text{-stat} = \dfrac{R^2_{ur} - R^2_{r}}{1 - R^2_{ur}} \cdot \dfrac{n-k-1}{q} \sim F_{q, n-k-1}$

$$F\text{-stat} = \frac{0.0189}{0.9811} \cdot \frac{110}{3} = 0.706 \sim F_{3,110}$$

and at $\alpha = 5\%$ the critical value is ≈ 3.07
(C_α)

Since $C_\alpha > F\text{-stat}$, we fail to reject H_0 and we conclude that the model does not exhibit heteroskedasticity at $\alpha = 5\%$.

Denne kolonne er forbeholdt sensor

This column is for external examiner

(b) In order to identify/estimate α_1, α_2 correctly we need the order- and rank condition to hold. The order condition states that we need at least one exogenous variable in the model we want to estimate, and the rank condition states that we need at least one ~~equati~~ exogenous and relevant variable in the other equation, this is because we want these variables in the other equation as 'shifters'.

Our problem here is that the models suffer from endogeneity bias, and we need to solve it. To show the bias, I insert (1) in (2) and solve for open.

Then $open = \beta_{20} + \alpha_2(\beta_{10} + \ldots + \beta_{12}\, oil + u_1) + \beta_{21} lpcinc + \ldots + u_2$.

$open(1-\alpha_1\alpha_2) = \beta_{20} + \alpha_2\beta_{10} + \alpha_2\beta_{11} lpcinc + \alpha_2\beta_{12}\, oil + \alpha_2 u_1$
$+ \beta_{20} + \beta_{21} lpcinc + \beta_{22} lland + u_2$

$\Rightarrow open = \pi_0 + \pi_1 lpcinc + \pi_2\, oil + \pi_3\, lland + \frac{\alpha_2 u_1}{1-\alpha_1\alpha_2} + u_2$

and we assume $cov(u_1, u_2) = 0$, $1-\alpha_1\alpha_2 \neq 0$.

Now we see that $cov(open, u_1) = E(open \cdot u_1) - \underbrace{E(u_1)}_{=0} \cdot E(open)$

$\Rightarrow \frac{\alpha_2 \sigma_{u_1}^2}{1-\alpha_1\alpha_2} \neq 0$ which proves endogeneity in open in model (1).

A similar result can be shown for inf, where

$cov(inf, u_2) = \frac{\alpha_1 \sigma_{u_2}^2}{1-\alpha_1\alpha_2} \neq 0.$

Kandidat nr. / Candidate no. 10046

Dato / Date 2015/19 Side / Page 8

Emnekode / Subject: SØK3001 Antall ark / Number of pages: 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

Now, to the point: We can test if the rank condition holds for both equations.

We now have two equations:

$open = \pi_0 + \pi_1 lpcinc + \pi_2 oil + \pi_3 lland + \varepsilon_0$

$inf = \theta_0 + \theta_1 lpcinc + \theta_2 lland + \theta_3 oil + \varepsilon_1$

In order to be able to identify α_1, we need lland to be significant in ~~explatns~~ explaining open. To identify α_2, we need oil to be significant in explaing inf. So we perform two tests:

(1) $H_0: \pi_3 = 0$ against $H_1: \pi_3 \neq 0$ at $\alpha = 5\%$.

(2) $H_0: \theta_3 = 0$ against $H_1: \theta_3 \neq 0$ at $\alpha = 5\%$.

~~(1) t-stat = 0.1398 / 7.33 = ... ~ t_{110}, and $|c_{\alpha/2}| = 1.98$, so we fail to reject Ho. Then it follows that~~

(1) $t\text{-stat} = \dfrac{-7.56}{0.818} = -9.24 \overset{\sim t_{110}}{,}$ and $|c_{\alpha/2}| = 1.98$

We see that $|t\text{-stat}| > |c_{\alpha/2}|$, so we can reject the null, and it follows that α_1 can be identified by the rank condition.

(2) $t\text{-stat} = \dfrac{-6.7}{9.7} = -0.69 \sim t_{110}$.

Here $|t\text{-stat}| < |c_{\alpha/2}| = 1.98$, so we fail to reject the null, and it follows that α_2 cannot be identified

November 2018

Kandidat nr. / Candidate no. 10046

Dato / Date 2015/19 Side / Page 9

Emnekode / Subject: SØK3001 Antall ark / Number of pages: 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

(c) In general, we wish that our IV is exogenous ($cov(z,u)=0$), and relevant ($cov(x,u) \neq 0$). Also, we want our IV to be as strong as possible, preferably such that its t-stat > 3.2 and F-stat > 10. The reason we want strong IVs ~~are~~ is important as we can have bias and distorted inference, especially in small samples. To see this, ~~to~~ the IV estimator has

$$plim(\hat{\beta}_1^{IV}) = \beta_1 + \frac{corr(z,u)}{corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x}$$, and if

$corr(z,u) \neq 0$ (but small, this is not unrealistic for small samples), then a low correlation between z, x (read: low t-or F-stat), will produce a large bias, so ideally we want to maximize $corr(x,z)$.

So testing for relevance is one strategy. Also, if the IV and OLS estimates are very different ~~E~~, then it might be a sign of endogeneity in the IV (as discussed above). If everything above holds, picking the IV that is closest to the OLS estimate could be another strategy.

Lastly, since land is a strictly positive variable, its log, lland, will ~~have~~ be closer to its t- and F statistics, so that could also be an argument, for lland. eliminates skew

Kandidat nr. / Candidate no. 10046

Dato / Date 2015/19 Side / Page 10

Emnekode / Subject: SØK 3001 Antall ark / Number of pages: 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

10

(d) If inf is exogenous in equation (2), then IV and 2SLS estimation would give the same results. So we can perform an endogeneity test for inf, if inf is exogenous we do not need IV/2SLS estimation.

If we do the first stage regression on $inf = \pi_0 + \pi_1 lpinc + \pi_2 oil + \pi_3 lland + v$ and collect its residuals $\hat{v}$, then we can expand equation (2) with

$open = \beta_{20} + \alpha_2 Inf + \beta_{21} lpcinc + \beta_{22} lland + \delta \hat{v} + u$

Here, endogeneity would imply that $\delta \neq 0$ since there are parts of ~~open~~ inf that explain open, which is not captured by the exogenous variables. So we let

$H_0: \delta = 0$ against $\delta \neq 0$ at $\alpha = 5\%$.

and perform a t-test on δ.

So ~~they are the~~ OLS and IV estimators are the same if $\delta = 0$.

NTNU

Kandidat nr. / Candidate no. 10046

Dato / Date 2015/19 Side / Page 11

Emnekode / Subject: SØK3001

Antall ark / Number of pages: 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

(e) The 2SLS estimators ~~for~~ are found by regressing open on all exogenous variables, so

$open = \beta_0 + \beta_1 oil + \beta_2 lpcinc + \beta_3 lland$

and then we use $\hat{open}$ on the FOCs for the OLS method:

$$\sum_{i=0}^{n} \hat{u} = 0 \quad (1)$$

$$\sum_{i=1}^{n} \hat{open}(y_i - \hat{\beta}_0 - \ldots) = 0 \quad (2)$$

which yields the $\hat{\beta}_i^{2SLS}$ coefficients.

If the $\hat{\beta}_1^{2SLS}$ coefficient is to be consistent, we want $plim(\hat{\beta}_1^{2SLS}) = \beta_1^{2SLS} + \frac{cov(\hat{open}, u)}{Var(\hat{open})}$

~~which holds as long as the IVs for open~~

which holds as long as the IVs for open are exogenous, meaning they are not correlated with u. I believe they are exogenous, so 2SLS should be consistent.

~~Regarding efficiency I would be a bit unsure. Heteroskedastic-robust If there are signs of heteroskedasticity then a low sample size can cause bias in the variance calculation, and if this is the case with n = 114, then no.~~

Regarding efficiency: From (a) we found no signs of heteroskedasticity, so the robust standard errors will not be efficient, also since n is small (114) robust se's can be biased.

NTNU

Kandidat nr. / *Candidate no.* 10046

Dato / *Date* 2015/19 Side / *Page* 12

Emnekode / *Subject:* SØK3001

Antall ark / *Number of pages:* 12

Denne kolonne er forbeholdt sensor

This column is for external examiner

#3

(a)

I believe they are using a difference-in-difference estimator to determine whether this new allocation of police forces has helped reduce crime.

The model would be

$$CarTheft_{it} = \beta_0 + \cancel{\beta_1 policeman_{it}} + \beta_2 NewPolice_{it} + v_{it} + a_i$$

which is a model using POLS.

The idea is that you want to compare the effects after and before the allocation of additional police forces.

So after − before $= \underbrace{\beta_0 + \beta_2 NewPolice_{it}}_{after} - \underbrace{(\beta_0 + \beta_2 NewPolice_{it})}_{before}$

but in 'before' we have that $NewPolice_{it} = 0$ as it was before any new police were deployed. So the DiD estimator would be $\beta_0 + \beta_2 - \beta_0 - \beta_2 \cdot 0 = \underline{\beta_2 NewPolice_{it}}$, which means that the effect after allocating new forces would be $\beta_2 NewPolice_{it}$.

(b) β_0 is the average value or numbers of cars stolen of car theft in Argentina.

β_2 is the change in car theft in location i at time t after an additional policeman in that ~~area~~ location i at time t. has been deployed

From the conclusion of the article I expect $\beta_2 < 0$, and possibly large,