Yulia Volkova

Problem Set 1

① NW do not alter each ρ_j individually

If we have $Var(\hat{\beta}) = \left(\frac{1}{T}\frac{\sigma^2_v}{(\sigma^2_x)^2}\right)\left(1+2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j\right)$

where $v_t = (x_t - \mu_x)u_t$

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}}{\sum^{T}\hat{v}_t^2}$$

NW suggest summing cross terms over M instead of T where M is a function of T but does not grow as fast.

$$Var(\hat{\beta}) = \left(\frac{1}{T}\frac{\sigma^2_v}{(\sigma^2_x)^2}\right)\left(1+2\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)\hat{\rho}_j\right)$$

So we will have flexible amount of ρs that depend on T. Every time we will be adding another T to our observations, we will not be adding another ρ_j and we will be able to avoid incidental parameters problem and achieve consistency.

② $y_{it} = \alpha_{it} + \beta x_{it} + \varepsilon_{it}$

If α_{it} is growing over time and is correlated with x_{it} then F.E. will not take care of all individual heterogeneity. Time-varying part of α_{it} will stay in the regression error and so $\hat{\beta}$ will not be unbiased. ***

Time-fixed effects also won't help because only $x_i = 1$ will have a time trend. So the trend is not common across entities and will not be

time constant part in the regressor → $X = \begin{bmatrix} \alpha_i & \vdots \\ \vdots & x_i \end{bmatrix}$ $\quad u_{it} = \begin{bmatrix} \lambda_t + \varepsilon_{it} \end{bmatrix}$ ← time varying part in composite error

***

taken away by time F.E. Trend is correlated with value of X and only average will be taken off. The trend will survive.

(3) Entity dummies fixed effects will not help when we are dealing with non stationarity. So of course FD estimation will be superior

FD: $\Delta y_t = \Delta X_t \beta + \Delta \varepsilon_t$

$\Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$

$\Delta \varepsilon_{t-1} = \varepsilon_{t-1} - \varepsilon_{t-2}$

FD will make X stationary.

Time F.E will help in case trend is common across all units. If this is the case time F.E will be superior to F.D since F.E preserve more information as when we first-difference the data we lose some information.

(4) Since there may be serial correlation within entities across time st. errors will be inconsistent even with state F.E 2 year F.E. If we do not cluster and our st. errors are including summation over T we will not achieve consistency. We need to cluster at state level to allow for serial correlation below this level as in Arellano. Then T will be mashed up inside the standard errors and averaging by N will give consistency

$$0 = \frac{1}{N} \sum_{i}^{N} \ddot{x}_i' \hat{u}_i \hat{u}_i' \ddot{x}_i$$

However, we will lose information as we are now using only N pieces of information.

(5) If serial correlation is negative, ρ_j is negative. Therefore, it can be that

$$\left(1 + 2\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)\hat{\rho}_j\right) < 0$$

if $\hat{\rho}_j$ is negative enough. Then $var(\hat{\beta})$ which is

$$\left(\frac{1}{T}\frac{\sigma'^2 v}{(\sigma^2 x)^2}\right)\left(1 + 2\sum_{j=1}^{M-1}\left(\frac{M-j}{M}\right)\hat{\rho}_j\right) < 0$$

when second term < 0.

(6) P-value associated with some state of the world θ_0 is the prob-ty of observing data Y at least as extreme as the data that was actually observed, y, when θ_0 is the true state of the world.

So p-value of 0.02 means that there is at most 2% prob-ty that candidates have equal support in South Africa today. We can also observe a more adverse draw.

Long Question

(1) $Y_{ijt} = \overset{dit+}{\beta}(takeup_{ijt}) + \delta_t + \delta_j + X_{ijt}\gamma + \varepsilon_{ijt}$

$\hat{\beta}$ will be the effect of one unit change in takeup on household income.

(2) Fixed effects will not be unbiased if there are interactions between d_{it} & x_{ijt}. The "whole" of alpha it will not be taken away by F.E. since now it is an interaction term with the Xs. So partly d_{it} will be in the error term

and cause bias for $\hat{\beta}$. This interaction makes marginal effect of x change.

So in our case not only do health concious families get more healthcare but it also means they get less use of additional increase in healthcare (vaccine).

(3) Yes, if vaccine takeup is indeed exogenous to household income we should be able to reasonably draw inference from coefficient on takeup ijt as the expected causal impact of taking the vaccine on a family's income. The extent to which the inference is reasonable will depend on the setting though. we may argue that external validity is not satisfied and expected causal impact of vaccine takeup on income is such only at the setting we are in.

If treatment is exogenous, then it is chosen without respect to how it might influence the outcome of interest.

(4) A disadvantage of this empirical strategy will be loss of precision/information. We move from 10.000 families to 20 communities. So now we only use across-communities variation for β. Before we could compare people in the same community but with different vaccination takeup.

The advantage is that on individual level the F.E. assumption is very unlikely to hold

$E[u_{ij}|d_{it}, x_{ijt}] = 0$ seems to be a strong assumption. Also on individual level we are likely to have SUTVA violations. With vaccinations case the no interference SUTVA assumption is highly unlikely. Treatment applied to some individuals in the community will have effect on other individuals as it will decrease their chances of getting malaria → lower health level → lower income. This is much less likely to happen when we focus on separate communities instead.

Also in real world people are more likely to be interested in an effect on community rather than individuals because vaccination policy is likely to be implemented on community level.

(6) Interaction term coefficient captures the difference in the slopes of regression lines between income and infection levels before and after the vaccine becomes available. So β is the marginal effect of vaccine on income given the specific infection level before introduction of the vaccine.

(7) Just estimate pre-post comparison. Did the outcome increase over the post-pre time period or not. It is unclear what the control group would be since everyone had vaccination at same time. All the treatment effect will be identified with time trend.

(8) For diff-in-diff to work we need parallel trends assumption. So if slopes of trend are different in locations with high & low infection rates which is a plausible thing to say. Places with higher levels of infection will be likely to have lower income growth over time. Thus growth of low-infection regions is not a good counterfactual. So the critic is right & we cannot make correct inference, get correct coefficient estimates from diff in diff.

(9) I would use Arellano st. errors since we have clustered data and positive serial correlation over time & correlation between people in clusters. Arellano would allow for it and give consistent st. errors.

I think one could also use NW for the over time serial correlation as this approach is also usually plausible to achive consistent st. errors.

(10) We can conclude that there is not enough evidence to say that vaccination effect on income does depend on pre-vaccination infection level.

Policy implementation would be to assign same treatment irrespective of prior infection level.

However we could have made type II error. The parallel trends assumption seems unlikely to hold so diff-in-diff results may be invalid. If for example the highly infected regions had no growth at all, then same average post-vaccination outcome implies that there was large impact on highly infected areas.

Our test has a low power.