Causal Inference
Problem Set 5
Due Wednesday May 23

Problem 1

In this problem, we work with data from: "Does Direct Democracy Hurt Immigrant Minorities? Evidence from Naturalization Decisions in Switzerland," by Jens Hainmueller and Dominik Hangartner. It is available at: http://www.stanford.edu/∼jhain/research.htm. The data you need are `Swiss_Panel_Long.dta`. Instead of the original 1400 municipalities, to ease the computational burden here we use 516 municipalities for which (a) there was a change in treatment status, and (b) full panels are available on the most important variable.

Take a quick look at the paper. In Switzerland, naturalization requests of immigrants are decided at the municipal level and municipalities use different types of institutions to decide on the naturalization applications. The main types of institutions are

- Direct democracy: Citizens meet in an assembly meeting and directly vote on the naturalization application by hand raising.
- Representative democracy: Naturalization applications are decided in the elected municipality council.

The main idea of the paper is to see if immigrants fare better under direct or representative democracy. In other words, the outcome of interest is the local naturalization rate and the key independent variable is the institution used to decide on who gets naturalized and who does not in each municipality. The panel data go from 1991 to 2009. Throughout the 1990s, many (80%) of municipalities used direct democracy to approve or reject naturalization applications. However in July 2003, the Swiss Federal Court ruled that secret ballot voting for naturalization referenda violates the Swiss Constitution (see the paper for more details). Following this, many municipalities switched over to a representative democratic system for approving applications following this decision (see figure 1). The dummy variable for treatment will simply code whether a municipality in a given year uses *direct democracy* versus representative democracy (note that in this analysis, we will not concern ourselves with the differences between representative democracy and appointed commissions which is also used in the paper; only very few municipalities use appointed commissions).

- *Year* – the year.
- *DirDem* – the causal variable of interest, takes the value 1 for a municipality if it used direct democracy that year (as of January 1). Takes value 0 for representative democracy.

- *nat_rate_ord* – the main outcome of interest: the naturalization rate in a given municipality in a given year (i.e. number of naturalizations in year t divided by the size of the group of eligible immigrants at the beginning of year t).
- *ortname* – municipality name
- *svpzero* – a time-varying measure of the electoral support for the SVP, the right wing party. This is used to proxy the anti-immigrant preferences in the local electorates in the municipalities.
- *svpcontszero* – a time-invariant measure of support for the SVP measured once in 1991.
- *svpconstzerogr* – same as *svpconstzero*, but binned into three levels: low, medium, and high SVP vote shares in 1991.
- *sprachreg* – a variable indicating the primary language in each municipality. It takes values G for German, I for Italian, and F for French.

(a) Estimate the model $Y_{it} = \alpha_0 + \alpha_1 \text{Direct Democracy}_{it} + \epsilon_{it}$, and report your estimates (both point estimates and standard errors, making sure to use a variance estimator that takes the likely clustering of errors into account). Under what assumptions will the estimate of α_1 be consistent for the parameter of the best linear predictor function? Under what assumptions will $\hat{\alpha}_1$ be consistent for the ATT?

```
> load("swissnat_PS7.Rdata")
> d <- d[-which(is.na(d$nat_rate_ord)==TRUE),]
>
> d$repdem <- 1-d$DirDem
>
> d <- plm.data(d, indexes = c("ort_name", "year"))
>
> vcovClusterX <- function(
+   model,
+   cluster,
+   dummies = FALSE
+ )
+ {
+   require(sandwich)
+   require(lmtest)
+   if(nrow(model.matrix(model))!=length(cluster)){
+     stop("check your data: cluster variable has different N than model")
+   }
+   M <- length(unique(cluster))
```

```
+   N <- length(cluster)
+   K <- model$rank
+   A <- 0
+   if(dummies == TRUE){
+     A <- M
+   }
+   if(M<50){
+     warning("Fewer than 50 clusters, variances may be unreliable (could try block boots
+   }
+   dfc <- (M/(M - 1)) * ((N - 1)/(N - K + A))
+   uj  <- apply(estfun(model), 2, function(x) tapply(x, cluster, sum));
+   rcse.cov <- dfc * sandwich(model, meat = crossprod(uj)/N)
+   return(rcse.cov)
+ }
>
> moda <- lm(nat_rate_ord ~ DirDem, data=d)
> coeftest(moda,vcov = vcovClusterX(moda, cluster=d$ort_name,dummies=FALSE))

t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  4.56428    0.18716  24.387 < 2.2e-16 ***
DirDem      -2.58100    0.18722 -13.786 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

$\hat{\alpha}_1$ will be consistent for the parameter of the best linear predictor function without further assumptions. For the ATT, the following assumptions suffice:

(1) democracy is randomly assigned

(2) SUTVA holds.

(b) Now include municipality fixed effects η_i, and report the estimates (again both point estimates and standard errors, making sure to use a variance estimator that takes the likely clustering of errors into account). How do the assumptions from (a) change to interpret $\hat{\alpha}_1$ to be a consistent estimate for the ATT of direct democracy on naturalization rates?

```
> modb <- lm(nat_rate_ord ~ DirDem + ort_name, data=d)
> coeftest(modb,vcov = vcovClusterX(modb, cluster=d$ort_name,dummies=TRUE))[1:4,]
                          Estimate   Std. Error       t value      Pr(>|t|)
```

```
(Intercept)               5.6345889 8.145422e-02  6.917492e+01  0.000000e+00
DirDem                   -2.8448125 1.719589e-01 -1.654356e+01  1.306237e-60
ort_nameAdligenswil       1.7831447 8.145422e-02  2.189137e+01 1.269891e-103
ort_nameAeschi B. Spiez  0.3014976 1.294502e-12  2.329063e+11  0.000000e+00
```

Now, we need the following assumptions:

(1) $E(\epsilon_{it}|DD_{it}, \eta_i) = 0$

(2) treatment effects are constant,

(3) SUTVA holds.

(c) Next, estimate $Y_{it} = \eta_i + \beta t + \alpha_1 \text{Direct Democracy}_{it} + \epsilon_{it}$, where t is a common linear time trend and report the estimates (again both point estimates and standard errors, making sure to use a variance estimator that takes the likely clustering of errors into account). How do the assumptions from (a) and (b) change to interpret $\hat{\alpha}_1$ to be a consistent estimate for the ATT of direct democracy on naturalization rates?

```
> d$time <- as.numeric(d$year)
>
> modc <- lm(nat_rate_ord ~ DirDem + time + ort_name, data=d)
> coeftest(modc,vcov = vcovClusterX(modc, cluster=d$ort_name,dummies=TRUE))[1:4,]
                      Estimate Std. Error   t value     Pr(>|t|)
(Intercept)          2.9162069 0.21313490 13.682447 3.312608e-42
DirDem              -1.0770477 0.21223268 -5.074844 3.953280e-07
time                 0.1881020 0.01377682 13.653508 4.891614e-42
ort_nameAdligenswil  0.9457824 0.10053127  9.407843 6.274469e-21
```

Now, we need the following assumptions:

(1) $E(\epsilon_{it}|DD_{it}, \eta_i, t) = 0$

(2) treatment effects are constant,

(3) SUTVA holds.

(d) Replicate Table 1 of the paper, columns (1) and (3). The "German Language" columns refer to a subsample including only the German speaking municipalities. How do the assumptions from (a), (b) and (c) change to interpret $\hat{\alpha}_1$ to be a consistent estimate for the ATT of direct democracy on naturalization rates?

```
> modcol1 <- lm(nat_rate_ord ~ DirDem + year + ort_name, data=d)
> coeftest(modcol1,vcov = vcovClusterX(modcol1, cluster=d$ort_name,dummies=TRUE))[1:4,]
               Estimate Std. Error    t value     Pr(>|t|)
```

```
(Intercept)  3.58785335  0.2323527 15.4414117 3.942003e-53
DirDem      -1.59323360  0.2596249 -6.1366745 8.771914e-10
year1992     0.05923506  0.1968381  0.3009329 7.634724e-01
year1993     0.37232378  0.2360133  1.5775543 1.147022e-01
>
> dGerman <- d[d$sprachreg=="G",]
> dGerman$ort_name <- as.character(dGerman$ort_name)
> modcol3 <- lm(nat_rate_ord ~ DirDem + year + ort_name, data=dGerman)
> coeftest(modcol3,vcov = vcovClusterX(modcol3, cluster=dGerman$ort_name,dummies=TRUE))[1
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept)  2.9363037  0.2895559 10.1407141 5.722889e-24
DirDem      -1.3320433  0.3181128 -4.1873302 2.863628e-05
year1992     0.1773431  0.2168172  0.8179384 4.134254e-01
year1993     0.3707840  0.2735976  1.3552166 1.754004e-01
```

$\hat{\alpha}$ will be consistent for the ATT under the slightly weaker assumption that (a) $E(\epsilon_{it}|DD_{it}, \eta_i, t) = 0$ *among German-speaking municipalities only*, (b) treatment effects are constant, and (c) SUTVA holds.

(e) Under what assumptions will the estimated standard errors of the coefficients in table 1 be consistent?

Variance of the error term is allowed to vary by observation, and to be correlated across observations within clusters. However, no correlation between errors across clusters needs to be assumed. These estimates for $s.\hat{e}.(\hat{\beta})$ will be consistent; hence a reasonably large number of clusters is needed for good coverage properties.

(f) Repeat the replication of table 1, columns (1) and (3), but show what results you get when you add unit-specific linear and linear-and-quadratic time trends.

Do you think it is a good idea to add unit-specific linear time trends? Unit-specific quadratic time trends? Why or why not?

```
> d$timesq <- (d$time)^2
> dGerman$timesq <- (dGerman$time)^2
>
> modcol1.ult <- lm(nat_rate_ord ~ DirDem + ort_name*time + year, data=d)
> coeftest(modcol1.ult, vcov = vcovClusterX(modcol1.ult,d$ort_name,dummies=T))[1:4,]
                         Estimate   Std. Error       t value    Pr(>|t|)
(Intercept)             0.2188394 3.422502e-01  6.394134e-01 0.522570796
DirDem                 -0.8956263 2.938257e-01 -3.048155e+00 0.002309379
```

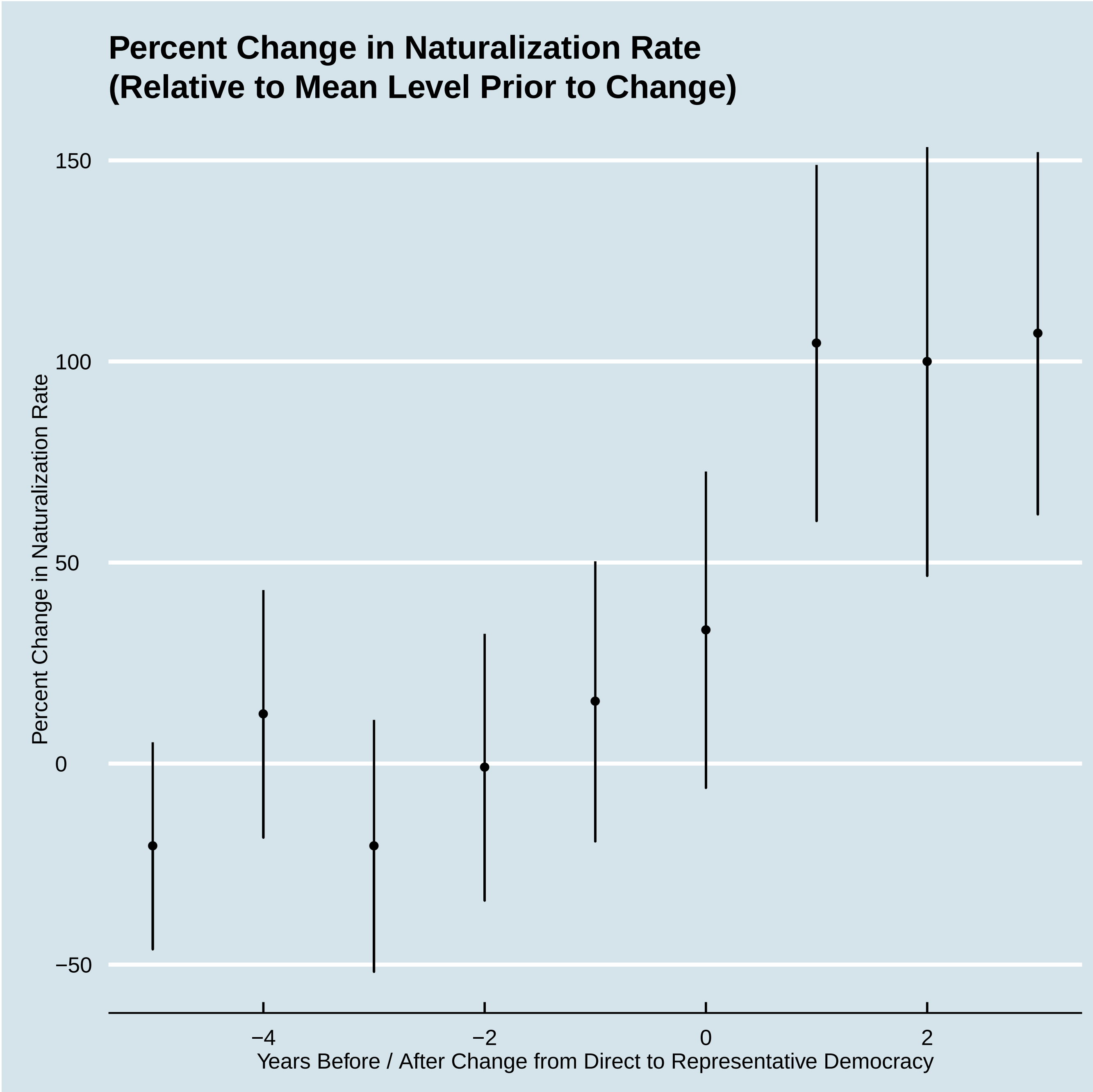
```
ort_nameAdligenswil      3.8330060 4.639353e-02  8.261942e+01 0.000000000
ort_nameAeschi B. Spiez  3.1384651 2.665984e-12  1.177226e+12 0.000000000
>
> modcol1.uqt <- lm(nat_rate_ord ~ DirDem + ort_name*time + ort_name*timesq + year,
 data=d)
> coeftest(modcol1.uqt, vcov = vcovClusterX(modcol1.uqt,d$ort_name,dummies=T))[1:4,]
                          Estimate   Std. Error       t value     Pr(>|t|)
(Intercept)              1.2471889 6.221986e-01  2.004487e+00 0.0450506000
DirDem                  -1.2217489 3.383323e-01 -3.611091e+00 0.0003067238
ort_nameAdligenswil     -0.6494546 1.686424e-01 -3.851075e+00 0.0001184914
ort_nameAeschi B. Spiez  1.7401259 8.477239e-11  2.052704e+10 0.0000000000
>
> modcol3.ult <- lm(nat_rate_ord ~ DirDem + ort_name*time + year,
 data=dGerman)
> coeftest(modcol3.ult, vcov = vcovClusterX(modcol3.ult,dGerman$ort_name,dummies=T))[1:4,]
                          Estimate   Std. Error       t value    Pr(>|t|)
(Intercept)              0.4292603 4.153152e-01  1.033577e+00 0.301378719
DirDem                  -0.9402673 3.617509e-01 -2.599212e+00 0.009368416
ort_nameAdligenswil      3.8259575 5.711856e-02  6.698274e+01 0.000000000
ort_nameAeschi B. Spiez  3.1384651 1.383734e-11  2.268114e+11 0.000000000
>
> modcol3.uqt <- lm(nat_rate_ord ~ DirDem + ort_name*time + ort_name*timesq + year,
 data=dGerman)
> coeftest(modcol3.uqt, vcov = vcovClusterX(modcol3.uqt,dGerman$ort_name,dummies=T))[1:4,]
                          Estimate   Std. Error       t value     Pr(>|t|)
(Intercept)              1.0770943 8.099370e-01  1.329850e+00 0.1836256436
DirDem                  -1.4004876 4.225888e-01 -3.314067e+00 0.0009257289
ort_nameAdligenswil     -0.7385473 2.106402e-01 -3.506202e+00 0.0004583521
ort_nameAeschi B. Spiez  1.7401259 1.360314e-11  1.279209e+11 0.0000000000
```

Unit specific time trends are definitely a good idea, because they allow for non-parallel time trends across different municipalities. There is little reason to believe that time trends/shocks (or time-varying omitted variables that could be partially taken into account by modeling time) are completely common across all municipalities, and thus modeling unit-specific time dynamics seems wise. Furthermore, because there are a substantial number of observations per municipality, quadratic unit-specific time trends also seem to be a good idea rather than simply linear unit-specific time trends. There is little reason to believe that any unit-specific time trend will be completely linear, and thus allowing for greater flexibility will do a better job capturing the effects of time. This would not be wise with too few observations per municipality, because it would lead to overfitting, but with the number in the data, quadratic modeling should capture true trends rather than chasing noise.

(g) Now perform the Autor Test and produce an Autor plot (as described on lecture slide 52, and shown on lecture slide 55). Note that this will not be an exact replication of the plot in the lecture slides because you are working with a subset of the data.

```
> # switches
> library(dplyr)
> d <- d %>% group_by(ort_name) %>% mutate(switcht = repdem - lag(repdem))
> d <- d %>% group_by(ort_name) %>% mutate(lag1  = lag(switcht,1))
> d <- d %>% group_by(ort_name) %>% mutate(lag2  = lag(switcht,2))
> d <- d %>% group_by(ort_name) %>% mutate(lag3  = lag(switcht,3))
>
>
> d <- ddply(
+   d, .(ort_name), transform,
+   lead1 = c( switcht[-1],NA )
+ )
> d <- ddply(
+   d, .(ort_name), transform,
+   lead2 = c( lead1[-1],NA )
+ )
> d <- ddply(
+   d, .(ort_name), transform,
+   lead3 = c( lead2[-1],NA )
+ )
> d <- ddply(
+   d, .(ort_name), transform,
+   lead4 = c( lead3[-1],NA )
+ )
> d <- ddply(
+   d, .(ort_name), transform,
+   lead5 = c( lead4[-1],NA )
+ )
>
> d$switcht[is.na(d$switcht)] <- 0
> d$lag1[is.na(d$lag1)] <- 0
```

```
> d$lag2[is.na(d$lag2)] <- 0
> d$lag3[is.na(d$lag3)] <- 0
>
> d$lag3 <- d$repdem - d$switcht - d$lag1 - d$lag2
>
> d$lead1[is.na(d$lead1)] <- 0
> d$lead2[is.na(d$lead2)] <- 0
> d$lead3[is.na(d$lead3)] <- 0
> d$lead4[is.na(d$lead4)] <- 0
> d$lead5[is.na(d$lead5)] <- 0
>
> d<-d[order(d$ort_name,d$year),]
>
> # autor
> mod_all <- plm(nat_rate_ord~lag3+lag2+lag1+switcht+lead1+lead2+lead3+lead4+lead5+
+                   year,data=d,model="within")
> holdem <- coeftest(mod_all, vcov=function(x) vcovHC(x, cluster="group", type="HC1"))
>
>
> meanbefore <- mean(d$nat_rate_ord[d$repdem==0])
>
> point <- (holdem[1:9,1]/meanbefore)*100
> lower <- (holdem[1:9,1] - 1.96*holdem[1:9,2])/meanbefore*100
> upper <- (holdem[1:9,1] + 1.96*holdem[1:9,2])/meanbefore*100
> yba <- rep(seq(from = 3, to = -5),3)
> type <- rep(c("point","lower","upper"),each=9)
>
> value <- c(point,lower,upper)
>
> plotdat <- data.frame(value,yba,type)
>
>
> outplot <- ggplot(plotdat, aes(x=yba,y=value)) + geom_line(aes(group=yba)) +
 geom_point(data=subset(plotdat,plotdat$type=="point")) +
+   theme_economist() +
xlab("Years Before / After Change from Direct to Representative Democracy") +
+   ylab("Percent Change in Naturalization Rate") +
ggtitle("Percent Change in Naturalization Rate (Relative to Mean Level Prior to Change)")
```



(h) We now want to examine how the effect of direct democracy changes depending on (some proxy for) the level of anti-immigrant attitudes in each municipality. What would be your hypothesis about how the effect of direct democracy varies as a function of SVP vote share (SVP is a right-wing populist political party in Switzerland)?

Higher SVP vote share should be associated with a larger effect (i.e. more negative) of direct democracy.

(i) Replicate Table B.7, columns 5 and 7. Note that you will use the coarsened version of time-invariant SVP vote share, *svpconstzerogr*.

```
> d$DirDem_med <- d$DirDem*(d$svpconstzerogr==1)
> d$DirDem_hi <- d$DirDem*(d$svpconstzerogr==2)
> modint1 <- lm(nat_rate_ord ~ DirDem + DirDem_med + DirDem_hi + ort_name + year, data=d)
> coeftest(modint1,vcov = vcovClusterX(modint1, cluster=d$ort_name,dummies=TRUE))[1:4,]
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  3.7353069  0.2503593 14.919787 9.243105e-50
DirDem      -0.6549871  0.4526485 -1.447010 1.479279e-01
DirDem_med  -0.8071366  0.4465788 -1.807378 7.073582e-02
DirDem_hi   -1.2564047  0.4662057 -2.694957 7.052458e-03

> dGerman$DirDem_med <- dGerman$DirDem*(dGerman$svpconstzerogr==1)
> dGerman$DirDem_hi <- dGerman$DirDem*(dGerman$svpconstzerogr==2)
> modint3 <- lm(nat_rate_ord ~ DirDem + DirDem_med + DirDem_hi + ort_name + year,
 data=dGerman)
> coeftest(modint3,vcov = vcovClusterX(modint3, cluster=dGerman$ort_name,dummies=TRUE))[1
              Estimate Std. Error    t value     Pr(>|t|)
(Intercept)  3.0784520  0.3046867 10.1036645 8.304331e-24
DirDem      -0.1618486  0.6284899 -0.2575199 7.967865e-01
DirDem_med  -1.0669292  0.6180199 -1.7263670 8.433374e-02
DirDem_hi   -1.4664463  0.6245010 -2.3481888 1.889779e-02
```

Problem 2

(a) Show that, with two periods $t \in 1, 2$, the fixed-effects model $y_{it} = x_{it}\beta + u_i + v_{it}$ (where u_i are a set of person-specific dummy variables) is equivalent to the first-differences model $\dot{y}_{it} = \dot{x}_{it}\beta + \dot{v}_{it}$, where $\dot{y}_{it} = y_{it} - y_{i,t-1}$.

$$\begin{aligned} y_{i,t} &= x_{i,t}\beta + u_i + v_{i,t} \\ y_{i,t} - y_{i,t-1} &= x_{i,t}\beta + u_i + v_{i,t} - y_{i,t-1} \\ \dot{y}_{i,t} &= x_{i,t}\beta + u_i + v_{i,t} - (x_{i,t-1}\beta + u_i + v_{i,t-1}) \\ &= (x_{i,t} - x_{i,t-1})\beta + (u_i - u_i) + (v_{i,t} - v_{i,t-1}) \\ &= \dot{x}_{i,t}\beta + 0 + \dot{v}_{i,t} \\ &= \dot{x}_{i,t}\beta + \dot{v}_{i,t} \end{aligned}$$

(b) Similarly, show that, with two periods, the fixed-effects model is equivalent to the *demeaned* model: $\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{v}_{it}$, where $\ddot{y}_{it} = y_{it} - \frac{1}{T}\sum y_{it} = y_{it} - \bar{y}_i$.

$$\begin{aligned} y_{i,t} &= x_{i,t}\beta + u_i + v_{i,t} \\ y_{i,t} - \frac{1}{T}\sum_t y_{i,t} &= x_{i,t}\beta + u_i + v_{i,t} - \frac{1}{T}\sum_t y_{i,t} \\ y_{i,t} - \bar{y}_i &= x_{i,t}\beta + u_i + v_{i,t} - \bar{y}_i \\ \ddot{y}_{i,t} &= x_{i,t}\beta + u_i + v_{i,t} - (\bar{x}_i\beta + \bar{u}_i + \bar{v}_i) \\ &= (x_{i,t} - \bar{x}_i)\beta + (u_i - \bar{u}_i) + (v_{i,t} - \bar{v}_i) \\ &= \ddot{x}_{i,t}\beta + 0 + \ddot{v}_{i,t} \\ &= \ddot{x}_{i,t}\beta + \ddot{v}_{i,t} \end{aligned}$$