ECON 710A - Problem Set 6

Alex von Hafften*

3/8/2021

1. Consider a random effects model with an intercept only: $Y_{it} = \mu_0 + \alpha_i + \varepsilon_{it}$ $(i = 1, ..., n, t = 1, ..., T_i)$ with $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT_i})'$ independent across i, $E[\alpha_i] = E[\varepsilon_i] = 0$, and

$$Var\begin{pmatrix}\alpha_i \\ \varepsilon_i\end{pmatrix} = \begin{bmatrix}\sigma_\alpha^2 & 0 \\ 0 & \sigma^2 I_{T_i}\end{bmatrix} \text{ where } \sigma^2 > 0$$

(i) Argue that the OLS estimator of μ_0 (the sample average), can be represented as $\hat{\mu}_{OLS} = \frac{\sum_{i=1}^n 1_i' Y_i}{\sum_{i=1}^n 1_i' 1_i}$ where $1_i = (1, ..., 1)' \in \mathbb{R}^{T_i}$ and $Y_i = (Y_{i1}, ..., Y_{iT_i})'$.

...

(ii) For a non-random instrument $Z_i \in \mathbb{R}^{T_i}$, consider the IV estimator $\hat{\mu}_{IV} = \frac{\sum_{i=1}^n Z_i' Y_i}{\sum_{i=1}^n Z_i' 1_i}$. Show that $Var(\hat{\mu}_{IV}) = \frac{\sum_{i=1}^n Z_i' \Omega_i Z_i}{(\sum_{i=1}^n Z_i' 1_i)^2}$ for some $\Omega_i = \Omega_i(\sigma_\alpha^2, \sigma^2, T_i)$ and find Ω_i.

...

(iii) Show that $Var(\hat{\mu}_{IV}) \geq (\sum_{i=1}^n 1_i' \Omega_i^{-1} 1_i)^{-1}$ and find an instrument $\tilde{Z}_i$ (possibly depending on Ω_i) such that $Var(\frac{\sum_{i=1}^n \tilde{Z}_i' y_i}{\sum_{i=1}^n \tilde{Z}_i' 1_i}) = (\sum_{i=1}^n 1_i' \Omega_i^{-1} 1_i)^{-1}$.

...

(iv) The IV estimator that uses $\tilde{Z}_i$ is often referred to as generalized least squares (GLS) and it is (weakly) more efficient than the OLS estimator. Is GLS strictly more efficient than OLS if $T_i = T$ for all i?

...

(v) Show that $\hat{\sigma}_i^2 = \frac{1}{T_i - 1} \sum_{t=1}^{T_i} (Y_{it} - \bar{Y}_i)^2$ where $\bar{Y}_i = \frac{1}{T_i} \sum_{i=1}^{T_i} Y_{it}$ has expectation equal to σ^2 and argue (informally) that $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_i{}^2$ is consistent for σ^2 as $n \to \infty$.

...

(vi) Show that $\hat{\sigma}_{\alpha,i}(\mu) = \frac{1}{T_i} \sum_{t=1}^{T_i} (Y_{it} - \mu)^2 - \hat{\sigma}_i^2$ has expectation equal to σ_α^2 when $\mu = \mu_0$ and argue (informally) that $\hat{\sigma}_\alpha^2 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{\alpha,i}^2(\hat{\mu}_{OLS})$ is consistent for $\hat{\sigma}_\alpha^2$ as $n \to \infty$.

...

(vii) If we let $\hat{\Omega}_i = \Omega_i(\hat{\sigma}_\alpha^2, \hat{\sigma}^2, T_i)$ and use this to construct a feasible version of $\tilde{Z}_i$, then we obtain the feasible GLS estimator. In the panel data context, this estimator is also referred to as the random effects estimator. The previous arguments imply that $\sqrt{n}\hat{\mu}_{FGLS}$ and $\sqrt{n}\hat{\mu}_{GLS}$ has the same asymptotic variances as $n \to \infty$. Propose a variance estimator $\hat{V}$ for which you expect (based on the previous questions) that $\hat{V}^{-1/2}(\hat{\mu}_{FGLS} - \mu_0) \to_d N(0, 1)$.

...

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Sarah Bass, Emily Case, Danny Edgel, and Katherine Kwok.

2. Consider a fixed effects regression model $Y_{it} = X_{it}\beta_0 + \alpha_i + \varepsilon_{it}$ $(i = 1, ..., n, t = 1, ..., T)$ where data is independent across i, X_{it} is independent of X_{is} for $s \neq t$ with $E[X_{it}] = 0$ and $E[X_{it}^2] = \sigma_x^2$ for all t, and strict exogeneity fails since $E[X_{is}\varepsilon_{it}] = \delta\sigma_x^2 1\{s = t+1\}$. Such failure can happen if the regressor is a response to a previous shock, e.g., $X_{it} = \delta\varepsilon_{i,t-1} + u_{it}$ for $t \geq 1$ and ε_{it} independent of u_{it}.

(i) Derive asymptotic biases of the fixed effects estimator and the first differences estimator as $n \to \infty$. [1]

Notice:

$$\begin{aligned}
E[X_{i,t}\varepsilon_{i,t}] &= \delta\sigma_x^2 1\{t = t+1\} \\
&= \delta\sigma_x^2(0) \\
&= 0 \\
E[\bar{X}_i\varepsilon_{i,t}] &= E\left[\frac{1}{T}\sum_{s=1}^{T} X_{i,s}\varepsilon_{i,t}\right] \\
&= \frac{1}{T}\sum_{s=1}^{T} E[X_{i,s}\varepsilon_{i,t}] \\
&= \frac{1}{T}\sum_{s=1}^{T} \delta\sigma_x^2 1\{s = t+1\} \\
&= \frac{(T-1)\delta\sigma_x^2}{T}
\end{aligned}$$

$$E[\bar{X}_i^2] = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E[X_{i,t}X_{i,s}] = \frac{1}{T^2}\sum_{t=1}^{T} E[X_{i,t}^2] = \frac{\sigma_x^2}{T}$$

From the hint, the numerator of the asymptotic bias of the fixed effects estimator is:

$$\begin{aligned}
E[\sum_{t=1}^{T}(X_{i,t} - \bar{X}_i)\varepsilon_{i,t}] &= \sum_{t=1}^{T} E[X_{i,t}\varepsilon_{i,t}] - \sum_{t=1}^{T} E[\bar{X}_i\varepsilon_{i,t}] \\
&= -\sum_{t=1}^{T}\frac{(T-1)\delta\sigma_x^2}{T} \\
&= -(T-1)\delta\sigma_x^2
\end{aligned}$$

The denominator is:

$$E[\sum_{t=1}^{T}(X_{i,t} - \bar{X}_i)^2] = \sum_{t=1}^{T} E[X_{i,t}^2] - 2\sum_{t=1}^{T} E[X_{i,t}\bar{X}_i] + \sum_{t=1}^{T} E[\bar{X}_i^2] = T\sigma_x^2 - 2\frac{\sigma_x^2}{T} + \frac{\sigma_x^2}{T} = T\sigma_x^2 - 2\frac{\sigma_x^2}{T} + \frac{\sigma_x^2}{T}$$

Thus asymptotic bias is:

$$\frac{E[\sum_{t=1}^{T}(X_{i,t} - \bar{X}_i)\varepsilon_{i,t}]}{E[\sum_{t=1}^{T}(X_{i,t} - \bar{X}_i)^2]} = \frac{\delta\sigma_x^2}{2T\sigma_x^2} = \frac{\delta}{2T}$$

[1] You may use without proof that $\hat{\beta}_{FE} \to_p \beta_0 + \frac{E[\sum_{t=1}^{T}(X_{i,t}-\bar{X}_i)\varepsilon_{i,t}]}{E[\sum_{t=1}^{T}(X_{i,t}-\bar{X}_i)^2]}$ and $\hat{\beta}_{FD} \to_p \beta_0 + \frac{E[\sum_{t=1}^{T}(X_{i,t}-X_{i,t-1})(\varepsilon_{it}-\varepsilon_{it-1})]}{E[\sum_{t=1}^{T}(X_{i,t}-X_{i,t-1})^2]}$

(ii) Is there a value of T so that the two asymptotic biases are identical?

...

3. Consider panel data $\{\{(Y_{it}, X_{it})'\}_{t=1}^{T}\}_{i=1}^{n}$ generated by the model $Y_{it} = X_{it}\beta_0 + \delta_t + \alpha_i + \varepsilon_{it}$, where $\{(Y_{it}, X_{it})'\}_{t=1}^{T}$ is i.i.d. across i, $T = 4$, and $(X_{i1}, ..., X_{i4})' = (0, 0, 1, 1)' \cdot 1\{\alpha_i > .6\}$. The error terms has an autoregressive structure $\varepsilon_{it} = \phi\varepsilon_{i,t-1} + u_{it}$ for $t \geq 1$, and α_i, ε_{i0}, and $u_{i1}, ..., u_{i4}$ are i.i.d. $N(0, 1)$. The parameters take the values $\beta_0 = \delta_2 = \delta_3 = \delta_4 = 1$ while δ_1 is normalized to zero and omitted from the model. We will vary the autoregressive parameter in $\{0, 0.8\}$ and consider sample sizes $n \in \{40, 70, 100\}$.

(i) In a statistical software of your choice, generate data, estimate $(\beta_0, \delta_2, \delta_3, \delta_4)$ using the FE estimator, and calculate both a heteroskedasticity robust variance estimate and a cluster robust variance estimate where the clustering is at the individual level. Use these variance estimators to construct two different 95% confidence intervals for β_0. Additionally, calculate the OLS estimator of the regression coefficients in the (misspecified) common intercepts model $Y_{it} = \alpha + X_{it}\beta_0 + \delta_t + \varepsilon_{it}$.

```
beta <- 1
delta_2 <- 1
delta_3 <- 1
delta_4 <- 1

phis <- c(0, 0.8)
sample_sizes <- c(40, 70, 100)

results <- NULL

for (phi in phis) {
  for (n in sample_sizes) {
    data <- tibble(i = rep(1:n, each = 4),
                   t = rep(1:4, times = n),
                   alpha = rep(rnorm(n), each = 4),
                   x = case_when(alpha > .6 & t %in% 3:4 ~ 1, TRUE ~ 0),
                   epsilon_0 = rep(rnorm(n), each = 4),
                   u = rnorm(n * 4),
                   delta_2 = case_when(t == 2 ~ delta_2, TRUE ~ 0),
                   delta_3 = case_when(t == 3 ~ delta_3, TRUE ~ 0),
                   delta_4 = case_when(t == 4 ~ delta_4, TRUE ~ 0)) %>%
      group_by(i)  %>%
      mutate(epsilon = case_when(t == 1 ~ phi * epsilon_0 + u),
             epsilon = case_when(t == 2 ~ phi * lag(epsilon) + u, TRUE ~ epsilon),
             epsilon = case_when(t == 3 ~ phi * lag(epsilon) + u, TRUE ~ epsilon),
             epsilon = case_when(t == 4 ~ phi * lag(epsilon) + u, TRUE ~ epsilon),
             x_bar = mean(x),
             x_d = x - x_bar) %>%
      ungroup() %>%
      mutate(y = x * beta + delta_2 + delta_3 + delta_4 + alpha + epsilon,
             y_d = y - mean(y),
             ones = 1)

    z <- data %>%
      select(x_d, delta_2, delta_3, delta_4) %>%
      as.matrix()

    x <- data %>%
      select(ones, x, delta_2, delta_3, delta_4) %>%
      as.matrix()

    y <- data$y
```

```r
    ols <- solve(t(x) %*% x) %*% (t(x) %*% y)
    fe <- c(NA, solve(t(z) %*% z) %*% (t(z) %*% y))

    y - x %*% fe

    results <- tibble(t = n,
           phi = phi,
           name = c("alpha", "beta", "delta_2", "delta_3", "delta_4"),
           ols = as.numeric(ols),
           fe = as.numeric(fe)) %>%
      bind_rows(results)
  }
}
```

t	phi	name	ols	fe
100	0.8	alpha	-0.06	NA
100	0.8	beta	2.60	0.84
100	0.8	delta_2	1.04	1.11
100	0.8	delta_3	0.53	1.15
100	0.8	delta_4	0.54	1.16
70	0.8	alpha	0.00	NA
70	0.8	beta	2.87	1.14
70	0.8	delta_2	0.98	1.13
70	0.8	delta_3	0.44	1.04
70	0.8	delta_4	0.41	1.01
40	0.8	alpha	-0.38	NA
40	0.8	beta	3.11	1.78
40	0.8	delta_2	1.10	0.96
40	0.8	delta_3	0.70	0.93
40	0.8	delta_4	0.84	1.07
100	0.0	alpha	0.09	NA
100	0.0	beta	2.61	0.58
100	0.0	delta_2	0.78	0.96
100	0.0	delta_3	0.44	1.22
100	0.0	delta_4	0.51	1.30
70	0.0	alpha	-0.04	NA
70	0.0	beta	2.75	0.57
70	0.0	delta_2	0.78	0.79
70	0.0	delta_3	0.94	1.32
70	0.0	delta_4	0.57	0.95
40	0.0	alpha	0.00	NA
40	0.0	beta	2.69	1.52
40	0.0	delta_2	0.94	1.16
40	0.0	delta_3	0.52	1.09
40	0.0	delta_4	0.40	0.98

(ii) Across 10000 simulated repetitions of the above, report the mean of the two point estimators for β_0 (OLS and FE) and the coverage rate for the two confidence intervals that rely on different variance estimators.

...

(iii) Discuss the results and relate them to the theory presented in lecture.

. . .