Econometrics II chearsheet

Eyayaw Beze

July 20, 2020

Assumptions of cross-section Regression:

(1) The data (y_i, x_i) are ***inid*** over i—stratified random sampling.
(2) The model is correctly specified . . . says: **linear in parameters**, **no measurement error, no omitted variable**.

Advantages of panel data (pd) against cross-sectional data:

1. Increased precision in estimation
2. Solves Omitted variable bias
3. Dynamics of individual behavior—lagged outcome variables.

Strict-exogeneity: $\mathrm{E}\left[\varepsilon_{it} \mid \alpha_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right] = 0, \quad t = 1, \ldots, T$ i.e., that ε_{it} has mean zero conditional on past, current and future values of $\mathbf{x}$.

POLS: most restrictive model, specifies **constant-coefficients, uses both t and i variations.** u_{it} is likely to be correlated over time for a given individual. Inconsistent if fixed effects model is true. Is appropriate if the constant-coefficients or RE models are appropriate, but panel-corrected standard errors and t-stats must be used for inference. However, $\hat{\boldsymbol{\beta}}_{\mathrm{POLS}}$ will be inefficient and a standard variance matrix is wrong unless additional assumptions of **homoskedasticity** and **no serial correlation** are maintained.

(a) $\mathrm{E}\left[u_{it}^2 \mathbf{x}_{it}' \mathbf{x}_{it}\right] = \sigma^2 \mathrm{E}\left[\mathbf{x}_{it}' \mathbf{x}_{it}\right]$

(b) $E\left[u_{it} u_{is} \mathbf{x}_{it}' \mathbf{x}_{is}\right] = \mathbf{0}, t \neq s$

BE: uses cross-section variation only. Is consistent if $\bar{x}_i$ are independent of the composite error. Inconsistent for fixed effects model as $Cov(\alpha_i, x_{it} \text{ hence } \bar{x}_i) \neq 0$

FE/FD: coefficients of time-invariant regressors are not identified.
The **incidental parameters problem** implies inconsistency of β_{LSDV} in short panels as only $N \to \infty$ since there are $N + (T-1) + K$ parameters with only NT observations—the number of parameters grows w/ the N. The FE estimator is efficient under the assumption of iid errors (FE.3). $\mathrm{E}\left[\varepsilon_i \varepsilon_i' \mid \mathbf{x}_i, \alpha_i]\right] = \sigma_u^2 I_T$.

FD is also efficient under FD.3: (a) $\mathrm{E}\left[e_{it}^2 \mathbf{x}_{it}' \mathbf{x}_{it}\right] = \sigma^2 \mathrm{E}\left[\mathbf{x}_{it}' \mathbf{x}_{it}\right]$ (b) $\mathrm{E}\left[e_{it} e_{is} \mathbf{x}_{it}' \mathbf{x}_{is}\right] = \mathbf{0}, t \neq s$ where $e = \Delta \varepsilon_{it}$, yielding the error structure $\varepsilon_{it} = \varepsilon_{i,t-1} + e_{it}$. FE.3 maintains that errors are serially uncorrelated, FD.3 maintains that errors follow a **random walk**.
with homoskedastic and serially uncorrelated ε_{it} **FE** is more efficient than **FD**. FD standard errors should be adjusted for the fact that the differenced residuals are serially correlated. POLS leaves α_i completely in the error term and RE only partly—less inconsistent. FE and FD estimators used to control for fixed effects typically remove both good and bad variation—susceptible to **attenuation bias** from measurement error.

- **RE**: inconsistent if the true model is the one with fixed effects. Treats α_i as unobserved random variable that is uncorrelated with the observed regressors $\mathbf{x}_i$. Because the omitted variable in a **RE** model is uncorrelated w/ included regressors there is no bias from ignoring it, it becomes part of the residuals. α_i *and* ϵ_{it} are **iid**—independent of regressors. The composite errors u_{it} will be serially correlated since $\text{Cov}\left[u_{it}, u_{is}\right] = \text{E}\left[\left(\alpha_i + \varepsilon_{it}\right)\left(\alpha_i + \varepsilon_{is}\right)\right] = \begin{cases} \sigma_\alpha^2, & t \neq s \\ \sigma_\alpha^2 + \sigma_\varepsilon^2, & t = s \end{cases}$ So POLS will be consistent but inefficient under the RE model. The most important consequence of **RE** is that the residuals for a given person are correlated across periods. Assumptions of the **RE** model: **linear CEF, homoskedasticity— equi-correlated errors**. The correlated between errors does not vary over time in the RE model. $\lambda \to 1 \; if \; T \to \infty$ or $\sigma_\alpha^2 > \sigma_\epsilon^2$: unobserved individual effects are important, the coefficients of time constant variables are difficult to estimate. $RE \to FE$.

correlated RE model–RE with addition of $\bar{x}_i$: models the correlated between α_i and $\boldsymbol{x}_i$. Adding the time average $\bar{x}_i$ and using RE is the same as subtracting the time average and using POLS.

Dynamic Panel Models: $y_{it} = \gamma y_{i,t-1} + \alpha_i + \varepsilon_{it}$ whenever we expect the dependent variable to exercise some kind of **persistence**/state dependence (SD) over time. Two sources: (1) **true SD**: $correlated[y_{i,t}, y_{i,t-1}] \simeq \gamma$ or $\alpha_i = 0$ (2) **unobserved heterogeneity** $correlated[y_{i,t}, y_{i,t-1}] \simeq \sigma_\epsilon^2/(\sigma_\alpha^2 + \sigma_\epsilon^2)$ or $\gamma = 0$—correlated of past and current outcomes due to unobserved individual effects. **All pd estimators are inconsistent if the regressors include lagged dependent variables**. The **FD** version of the dynamic model is inconsistent but we can use two-period lags as an instrument for $correlated[y_{i,t-1}, y_{i,t-2}]$ **Arellano-Bond estimator**, but errors be serially uncorrelated. **If the lagged dependent variables model is the correct one and you use FE, estimation of positive treatment effects tends to be too large. If FE is the correct model and you use lagged dependent variables your estimates tend to be too small**.

Mixed-Linear-Models: $y_{it} = \mathbf{z}'_{it}\boldsymbol{\beta} + \mathbf{w}'_{it}\boldsymbol{\alpha}_i + \varepsilon_{it}$ where $\mathbf{z}_{it}$ includes an intercept, $\mathbf{w}_{it}$ is a vector of observable characteristics, $\boldsymbol{\beta}$ is a fixed parameter vector and α_i is a random zero-mean vector.

Clustered standard errors—Abadie et.al 2007: motivation is that unobserved components of outcomes for units within clusters are correlated. Clustering is a **sampling design** issue when the sampling follows a 2-stage process, a subset of clusters from a population of clusters, units from sampled clusters. In this case the clustering adjustment is justified by the fact that there are clusters in the population that we do not see in the sample. An **experimental design** issue, when clusters of units, rather than units, are assigned to a treatment.

Clustering Matters Only if the residuals and the regressors are both correlated Within clusters—not necessarily; what matters is the within-cluster correlated of the product of the residuals & the regressors.

$Var_{clu} \geq Var_{het}$ due to the addition of terms when $i \neq j$, for $i, j \in c$.

The amount of increase is larger, the:
1. more positively correlated the regressors are across observations in the same cluster (via $x_i x_j$) 2. more positively correlated are the errors in the same clusters (via $u_i u_j$). 3. more observations are in the same cluster. **Aggregated regressor** $\uparrow$CRSE even if there is low within-cluster error correlated, because the regressor of interest is perfectly correlated within clusters.

The researcher should assess whether the sampling process is clustered or not, and whether the assignment mechanism is clustered. If the answer to both is no, one should not adjust the standard errors for clustering, irrespective of whether such an adjustment would change the standard errors. In case of fixed effects, we need to cluster if either the sampling or assignment to treatment was clustered. However, cluster adjustments will only make an adjustment with fixed effects if there is heterogeneity in treatment effects. So, if there is no heterogeneity in the treatment effects, one need not adjust standard errors for clustering once fixed effects are included. *FE will generally not completely control for within-cluster error correlated, so still use CRVE. Because FE may take out correlated e.g. within schools, but additional correlated within classrooms possible; or serial correlated in the error within cluster (e.g. same persons).* What to cluster over? Larger and fewer clusters have less bias at the cost of more variability. If clusters are **nested** rather use the broader cluster.

DiD: is a version of FE/FD estimation using aggregated data.

$$\hat{\tau}_{DID} = \left[\bar{Y}_1^T - \bar{Y}_0^T\right] - \left[\bar{Y}_1^C - \bar{Y}_0^C\right]$$

- **Stable Unit treatment Value Assumption (SUTVA)**, implies that treatment does not indirectly affect outcomes of the untreated—**no peer or GE effects**.

- **CIA** states that conditional on x, the outcomes are independent of treatment, $y_0, y_1 \perp \mathbf{D} \mid \mathbf{x}$.

- **Common Trend**: outcome trends would be the same in both groups in the absence of treatment. treatment induces a deviation from this CT. Additional covariets make the CT assumption more credible.

- **Common Support**: overlap between treated & untreated subsamples $0 < \Pr[\mathbf{D} = 1 \mid \mathbf{x}]$.

- **CMA/NEPT** $E\left[y_0 \mid \mathbf{D} = 1, \mathbf{x}\right] = E\left[y_0 \mid \mathbf{D} = 0, \mathbf{x}\right] = E\left[y_0 \mid \mathbf{x}\right]$, implies y_0 does not determine participation. No anticipation effects from being treated.

Bertrand et al.2004, 3 factors make serial correlated an especially important issue in the DID context: often long-timeseries used, dependent variable often highly positively serially correlated, treatment variables changes very little within a state over time (often intervention stays on).

IV: Solves endogen-and-simultaneity issues in OLS.

$y = \mathbf{x}'\boldsymbol{\beta} + u$; where $\mathrm{E}[\mathbf{x}u] \neq 0$, $\boldsymbol{\beta_{OLS}}$ is inconsistent.

Assumptions: Identification of β_{IV}

(IV.1): **Validity**/predeterminedness: $\mathrm{E}[\mathbf{z}u] = 0$; **(IV.2)**: **Relevance**. (a) rank $E\left[\mathbf{zz}'\right] = L$; (b) rank $E\left[\mathbf{z}x'\right] = K$; where $L \geqslant K$. And, we have $\widehat{\beta}_{IV} = [\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{Z}'y]$ $\widehat{\beta}_{2SLS} = \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}$ where $\hat{\mathbf{X}} = \mathbf{P_Z}\mathbf{X} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$. $\widehat{\beta}_{2SLS} = \widehat{\beta}_{IV}$; $if\ L = K$.
Homogeneous/Heterogenoues(LATE IV Estimator) treatment effects
For a simple model $Y_i = \alpha + \rho_i D_i + \epsilon_i$; where treatment $D_i = \pi_0 + \pi_{1i}z_i + \zeta_i$

Assumptions:

1. **(Independence)**: z_i is as good as randomly assigned; it is independent of the vector of potential outcomes and treatment assignments. $\{Y_i(D_{1i}, 1), Y_{0i}(D_{0i}, 0), D_{1i}, D_{0i}\} \perp\!\!\!\perp Z_i$. Independence implies that the **first-stage** is the average causal effect of Z_i on D_i

$$E[D_i \mid z_i = 1] - E[D_i \mid z_i = 0] =$$
$$E[D_{1i} \mid z_i = 1] - E[D_{0i} \mid z_i = 0] = E[D_{1i} - D_{0i}]$$

   Independence is also **sufficient** for a causal interpretation of the **reduced form**—the causal effect of the instrument on the dependent variable. Specifically, $E[Y_i \mid z_i = 1] - E[Y_i \mid Z_i = 0] = E[Y_i(D_{1i}, 1) - Y_i(D_{0i}, 0)]$

2. **Exclusion Restriction (ER)** $Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di}$ for $d = 0, 1$. instrument operates through a single known causal channel. In homogeneity, ER is expressed by $E[z_i\varepsilon_i] = 0$.

3. **(First-stage)** $E[D_{1i} - D_{0i}] \neq 0$

4. **(Monotonicity)** $D_{1i} - D_{0i} \geq 0$, or vice versa. $\rightarrow E[D_{1i} - D_{0i}] = P(D_{1i} > D_{0i})$, means that while the instrument may have no effect on some people, all of those who are affected are affected in the same way. Then,

$$\frac{E[Y_i \mid z_i = 1] - E[Y_i \mid z_i = 0]}{E[D_i \mid z_i = 1] - E[D_i \mid z_i = 0]}$$

The ER is distinct from the claim that the instrument is (as good as) randomly assigned. Rather, it is a claim about a unique channel for causal effects of the instrument. A failure of **monotonicity** means the instrument pushes some people into treatment while pushing others out—(**defiers**).

- **Compliers**. The subpop w/ $d_{1i} = 1$ & $d_{0i} = 0$:
- **Always-takers (AT)**. The subpop w/ $d_{1i} = d_{0i} = 1$:
- **Never-takers (NT)**. The subpop w/ $d_{1i} = d_{0i} = 0$:

without adding further assumptions (e.g., constant causal effects), LATE is not informative about effects on NT & AT because, by definition, treatment status for these two groups is unchanged by the instrument (random assignment).

LATE is the effect of treatment on the **population of compliers**. The average causal effect on compliers is not usually the same as the ATT. The treated consist of either AT or compliers with the **instrument switched on**.

$$\overbrace{E[Y_{1i} - Y_{0i} \mid D_i = 1]}^{\text{effect on the treated}} = \overbrace{E[Y_{1i} - Y_{0i} \mid D_{0i} = 1]}^{\text{effect on AT}} P[D_{0i} = 1 \mid D_i = 1]$$
$$+ \underbrace{E[Y_{1i} - Y_{0i} \mid D_{1i} > D_{0i}]}_{\text{effect on compliers}} P[D_{1i} > D_{0i}, z_i = 1 \mid D_i = 1]$$

The **ATT** (ATU) is a weighted average. of effects on **AT & compliers** (NT and compliers). population ATE = a weighted average of effects on compliers, AT, & NT.
The size of a compliers group is the Wald 1st-stage, $P(D_{i1} > D_{i0}) = E[D_{1i} \mid z_i = 1] - E[D_{0i} \mid z_i = 0]$. The proportion of the treated who are compliers is given by the 1st-stage, times the probability that the instrument is switched on $P[z_i = 1]$, divided by the prop treated $P[D_i = 1]$.
Staiger & Stock (1997) F < 10 may indicate problems with finite sample bias—weak instrument. IV may be more inconsistent than OLS when the correlated between the instrument and the regressor is low. IV **less precise** than OLS. $variable(\hat{\beta}_{IV}) \geq variable(\hat{\beta}_{OLS})$ unless $correlated(x, z) = 1$.

LPM Logit Probit

For general binary dependent variable model $p(\mathbf{x}) = \Pr(y = 1 \mid \mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}$
Marginal effects of a $x_j : \frac{\partial \Pr(y=1 \mid \mathbf{x})}{\partial x_j} = F'(x'\beta)\beta_j$

- Ignores discreteness & will treat the dependent variable as continuous.
- Does not constrain predicted probabilities to lie between 0 & 1.
- Does not take into account that outcomes might not be naturally ordered
- (+) Interest is in analyzing partial effects averaged over the distribution of x.
- (+) The x_j take on only a few values & the model is saturated. Thus, the CEF is (close to) a linear function of the regressors.
- For latent index model $z : F(z) = z; F(z) = \Phi(z); F(z) = 1/(1 + exp(-z))$
- The MLE is consistent if the CEF is correctly specified: $E[y|x] = F(x'\beta)$
- Identification of the single-index model requires a restriction of the variable (u) to secure uniqueness of β. In the probit model the error variance is set to one.
- Also the mean of the error distribution needs to be normalized: usually to zero.

For non-linear models, where

$F'\left(\mathbf{x}'\boldsymbol{\beta}\right) \neq c$ the me will vary with the evaluation point x and the choice of $F(\cdot)$:

1. (AME) : $N^{-1} \sum_i F'\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right) \widehat{\beta}_j$

2. The average of the regressors (MEM) : $F'\left(\overline{\mathbf{x}}'\widehat{\beta}\right) \widehat{\beta}_j$

3. The marginal effect at a representative point $\mathbf{x}_1$(MER) : $F'\left(\mathbf{x}_1'\widehat{\boldsymbol{\beta}}\right) \widehat{\beta}_j$

Threats to Internal Validity

(1.) Failure to Randomize: If the subjects are not randomly assigned to the treatment group. (2.) Failure to Follow the Treatment Protocol–defiers (3.) Attrition: If subjects systematically drop out of the study. (4.) Experimental effects: If subjects know that they are in an experiment. (5.) Small Sample Sizes

Threats to External Validity

(1.) Unrepresentative Sample, and Program or Policy (2.) GE effects: If market and/or environmental conditions cannot be kept constant when an internally valid program is implemented broadly, external validity may be doubtful.

RDD

RD methods exploit precise knowledge of the rules determining assignment to treatment around a threshold val of a variable. **SRDD**: treatment status is **deterministic** and a

discontinuous fun of covariate. We do not observe both treated and untreated values of x_i for individuals, thus validity of RD relies on our willingness to extrapolate in a neighborhood of c. model.eq: $Y_i = \alpha + \beta x_i + \tau D_i + \eta_i$ **FRDD** exploits discontinuities in the prob or expected val of treatment conditional on a covariate. Thus, the discontinuity becomes an IV for treatment and we use 2SLS.

Matching estimators are normally used when 1. The interest is in the ATT rather than the ATE. 2. Randomization into treatment is not possible. 3. There is a large pool of potential control subjects.