Recitation 6: Estimating Time Series, POLS, RE

Seung-hun Lee

Columbia University
Introduction to Econometrics II Recitation

February 28th, 2022

Time series regression

Building blocks for time series regressions

- Ergodicity theorem: For this, we want a sequence $\{Y_t\}$ that is strictly stationary and ergodic. Then if $E[Y_t] < \infty$, we have $\frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{p} E[Y_t]$
- White noise: An error e_t satisfying $E[e_t] = 0$, $var[e_t] = \sigma^2$ and $\gamma(k) = 0 \; \forall k \neq 0$
- Linear projection: Y_t can be projected onto the set of the history of the value of Y's up until period $t-1$. Such projection is unique and so is the resulting projection error. Specifically, we can write $Y_t = \mathbb{P}_{t-1}(Y_t) + e_t$. The resulting e_t is serially uncorrelated (and thus WN) and $\mathbb{P}_{t-1}(Y_t)$ is covariance stationary provided that $\{Y_t\}$ is.
- CLT: Let $\{y_t\}$ come from a dependent data satisfying strict stationarity and ergodicity. Specifically, let $y_t = \mu + e_t$ with e_t being a white noise. then, as $T \to \infty$, the following holds, $\sqrt{T}\frac{\bar{y}-\mu}{\omega} \sim N(0,1)$ where $\omega^2 = \sum_{j=-\infty}^{\infty} \gamma(j) = \gamma(0) + 2\sum_{j=1}^{\infty} \gamma(j)$

Wold representation: Stationary process has linear representation

- If we have a process $\{Y_t\}$ that is covariance stationary, we can represent this process as an infinite linear function of the projection errors
- Y_t can be written as a linear function of the white noise errors and a deterministic part

$$Y_t = \mu_t + \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

  If $\mu_t = \mu$, this is purely deterministic.
- This works for autoregressive and moving average processes

MA(1): Shortest memory process

- MA(1) model has the following form

$$y_t = \mu + e_t + \theta e_{t-1},\ e_t \sim (0, \sigma^2) \text{ is WN } \forall t \text{ (not necessarily normal)}$$

- We can calculate the key moment conditions as follows

$$\begin{aligned}
E[y_t] &= \mu + E[e_t] + \theta E[e_{t-1}] = \mu \\
\gamma(0) &= var(y_t) = var(\mu + e_t + \theta e_{t-1}) \\
&= var(e_t + \theta e_{t-1}) \ (\because \text{ constants do not affect dispersion}) \\
&= var(e_t) + \theta^2 var(e_{t-1}) = (1 + \theta^2)\sigma^2 \\
\gamma(1) &= cov(y_t, y_{t-1}) = cov(\mu + e_t + \theta e_{t-1}, \mu + e_{t-1} + \theta e_{t-2}) \\
&= \theta cov(e_{t-1}, e_{t-1}) = \theta\sigma^2 \\
\gamma(2) &= cov(y_t, y_{t-2}) \\
&= cov(\mu + e_t + \theta e_{t-1}, \mu + e_{t-2} + \theta e_{t-3}) = 0
\end{aligned}$$

MA(q): Short-memory property is generalizable

- For any $k > q$, the impact of the shock vanishes

$$y_t = \mu + \sum_{j=0}^{q} \theta_j e_{t-j}, \ \ e_t \sim (0, \sigma^2) \text{ is WN } \forall t$$

$$E[y_t] = \mu$$

$$\gamma(0) = var\left(\sum_{j=0}^{q} \theta_j e_{t-j}\right) = \left(\sum_{j=0}^{q} \theta_j\right) \sigma^2$$

$$\gamma(k) = cov(y_t, y_{t-k}) = cov\left(\mu + \sum_{j=0}^{q} \theta_j e_{t-j}, \mu + \sum_{j=0}^{q} \theta_j e_{t-k-j}\right)$$

$$= cov\left(\theta_k e_{t-k} + \ldots + \theta_q e_{t-q}, \theta_0 e_{t-k} + \theta_{q-k} e_{t-q}\right) = \left(\sum_{j=0}^{q-k} \theta_{j+k}\theta_j\right) \sigma^2$$

$$\gamma(k) = 0 \ \forall k > q$$

MA(q) can teach us about dynamic causal effect of shocks

- With white noise errors, dynamic causal effects can be cleanly analyzed

$$y_t = \sum_{j=0}^{q} \theta_j e_{t-j}$$

- We can find the impact of e_{t-j} with $\frac{\partial y_t}{\partial e_{t-j}} = \theta_j$.
- The interpretation of θ_j is the dynamic multiplier of e_{t-j} onto y_t.

AR(p) process is much more perserverant

- AR(p) has a Wold representation
- If we take $t \to \infty$, we can write y_t as a infinite sum of the white noise errors.

$$y_t = e_t + \alpha e_{t-1} + \alpha^2 e_{t-2} + \ldots = \sum_{j=0}^{\infty} \alpha^j e_{t-j}$$

- We effectively have MA(∞) process - a process that takes very long to die out.

How persistent is the effect?

- This is represented in these moment conditions.

$$\gamma(0) = cov\left(\sum_{j=0}^{\infty}\alpha^j e_{t-j}, \sum_{j=0}^{\infty}\alpha^j e_{t-j}\right) = \frac{\sigma^2}{1-\alpha^2}$$

$$\gamma(1) = cov\left(\sum_{j=0}^{\infty}\alpha^j e_{t-j}, \sum_{j=1}^{\infty}\alpha^{j-1} e_{t-j}\right) = \frac{\alpha\sigma^2}{1-\alpha^2}$$

$$\gamma(2) = cov\left(\sum_{j=0}^{\infty}\alpha^j e_{t-j}, \sum_{j=2}^{\infty}\alpha^{j-2} e_{t-j}\right) = \frac{\alpha^2\sigma^2}{1-\alpha^2}$$

$$\gamma(k) = \frac{\alpha^k\sigma^2}{1-\alpha^2}$$

- As long as $|\alpha| < 1$, this process does approximate to 0. Depending on the value of α the process can die off quickly, last long, or oscillate around 0.

Time series OLS is consistent but not unbiased

- We want to estimate α in $y_t = \alpha_1 y_{t-1} + e_t \;\; (e_t \sim (0, \sigma^2), |\alpha| < 1)$
- OLS estimate of α_1 is $\hat{\alpha}_1 = (X_t' X_t)^{-1}(X_t' y_t)$ where $X_t = [y_{t-1}]$
- Consistency? $X_t (= y_{t-1})$ and e_t and $X_t' e_t$ are ergodic stationary. If mean of y_t is finite, sample mean of $y_{t-1} e_t$ converges to 0. (Ergodicity theorem)
- Bias? Strict exogeneity $E[e_t | X_1, ..., X_T] = 0$. is no longer guaranteed.

$$
\begin{aligned}
E[e_t X_{t-1}] &= E[e_t y_{t-2}] = 0 \\
E[e_t X_t] &= E[e_t y_{t-1}] = 0 \\
E[e_t X_{t+1}] &= E[e_t y_t] = \sigma^2 (\neq 0)
\end{aligned}
$$

- Difference? Contemporary uncorrelatdness vs strict exogeneity

HAC standard errors are the standard

- The error structure needs to take time dependence!
- HAC: Has this variance structure

$$\Omega = \sum_{j=-\infty}^{\infty} \Gamma_g(j) = \Gamma_g(0) + \sum_{j=1}^{\infty} \left(\Gamma_g(j) + \Gamma_g(j)'\right)$$

  where g is the moment condition $E[X_t e_t] = 0$ and $\Gamma_g(k) = E[g_t g_{t-k}]$
- By ergodicity $\Gamma_g(k)$ is eventually zero $\rightarrow$How much weight on long term variances?
- Newey & West: Proposes ($M = 0.75 \times T^{1/3}$)

$$\widehat{\Omega} = \widehat{\Gamma}_g(0) + \sum_{j=1}^{M} \left(1 - \frac{k}{M+1}\right) \left(\widehat{\Gamma}_g(j) + \widehat{\Gamma}_g(j)'\right)$$

How about GLS?

- Recall that $D = E[ee'|X]$ can be represented by $D^{-1} = PP'$
- In time series, with AR(1) error term $e_t = \rho e_{t-1} + u_t$ (u_t is WN)

$$D = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & .... \\ \gamma(1) & \gamma(0) & \gamma(1) & .... \\ .... & ... & ... & .... \\ \gamma(T-1) & ... & ... & \gamma(0) \end{pmatrix} = \frac{\gamma(0)}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & .... \\ \rho & 1 & \rho & .... \\ .... & ... & ... & .... \\ \rho^{T-1} & ... & ... & 1 \end{pmatrix}, \quad P' = \frac{1}{\sigma_u} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & ... \\ -\rho & 1 & 0 & .... \\ .... & ... & ... & .... \\ 0 & ... & -\rho & 1 \end{pmatrix}$$

- Effectively, we transform (or quasi-difference) the DGP $y_t = x_t'\beta + e_t$ by

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + \underbrace{e_t - \rho e_{t-1}}_{u_t \rightarrow \text{spherical!}}$$

- we need to check whether a stronger condition $E[(x_t - \rho x_{t-1})u_t] = 0$ holds.

Model selection: How many lags?

- **Information criterion**: Gives the number that sums the measure of fitness of the model and the penalty for a complex model.
- They are calculated as

$$IC(k) = \underbrace{\log \hat{\sigma}^2}_{\text{Penalty for lack of fit}} + \underbrace{k\frac{C_T}{T}}_{\text{Penalty for complex model}}$$

where there are k total regressors

- Akaike information criterion: $C_t = 2$
- Bayesian information criterion: $C_T = \log T$

Panel Regression

Panel deals with entity/time-fixed unobserved factors

- Let y and $x = (x_1, x_2, ..., x_k)$ be the observable factors.
- α: Unobservable random variable incorporated into the DGP additively.
- Then, we can write $E[y|x, \alpha]$ as

$$E[y|x, \alpha] = x'\beta + \alpha$$

- α independent from x? β is consistently estimable.
- α correlated with x? Still a chance if α fixed across time/entities and have panel data
- Most discussion of panel data in class will focus on α representing an unobservable trait inherent for individuals or a common trend in a particular period

Setting up the panel regression

- Our data generating process is

$$y_{it} = x_{it}'\beta + v_{it} \iff y_i = X_i\beta + v_i$$

  where $x_{it} \in \mathbb{R}^k$, $y_{it} \in \mathbb{R}$, and $y_i \in \mathbb{R}^T$, $X_i \in \mathbb{R}^{T\times k}$
- v_{it}: ‘composite’ error term
  - One-way FE: $v_{it} = c_i + e_{it}$
  - Two-way FE (TWFE): $v_{it} = c_i + \delta_t + e_{it}$
  - Interacted FE (IFE): $v_{it} = \lambda_i f_t + e_{it}$
- The assumption on c_i determines the method of panel regression we will use
- The assumptions on e_{it} the degree of exogeneity we are willing to assume.

Types of panel regression

- There are three types of estimators we study in (static) panel regression, determined by assumptions on $E[c_i|X_i]$

POLS, RE, FE

- Is c_i constant $\forall i$? Then **pooled OLS (POLS) estimator** is consistent and efficient
- Is c_i latent (in that it is unobserved and possibly different for each i) but uncorrelated with X_i? Then POLS is consistent but inefficient. We use **random effects (RE) GLS estimator** here
- Is c_i latent and possibly correlated with X_i? Then POLS and RE are inconsistent. We use **fixed effects (FE) estimator** - which are within estimation (WE), least squares dummy variables (LSDV), and in case we have $T = 2$, first-differencing (FD)

Types of exogeneity

- They have important bearings on consistency/unbiasedness

Strict and sequential exogeneity (predetermined regressors)

- **Strict exogeneity**: $E[e_{it}|x_{i1}, ..., x_{iT}, c_i] = 0$. Or that the past, current, or future values of x_{it} cannot predict e_{it}
  - Weaker version: $E[x_{is}e_{it}] = 0$ for $s = 1, ..., T$. This implies that e_{it} cannot be predicted by x_{is}
- **Sequential exogeneity**: $E[e_{it}|x_{i1}, ..., x_{it}, c_i] = 0$ for each t. Past and present value of x_{it} cannot predict e_{it}. It is silent on whether future values of x_{it} predicts e_{it}

Pooled OLS: c_i is a constant and just another overall intercept!

- The data generating process is written

$$y_{it} = x_{it}'\beta + c + e_{it}$$

- If we assume that $E[x_{it}e_{it}] = 0$, we end up with the following pooled OLS estimate

$$\begin{aligned}\hat{\beta}_{POLS} &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}y_{it}\right) \\ &= \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'y_i\right)\end{aligned}$$

Pooled OLS properties are much like OLS

- We can find unbiasedness, consistency, variance, and asymptotics

$$E[\hat{\beta}_{POLS}|X_i] = \beta + \left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}E[e_{it}|x_{i1},....,x_{iT},c]\right) = \beta$$

$$\text{plim } \hat{\beta}_{OLS} = \beta + \text{plim}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}x_{it}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}e_{it}\right) = \beta$$

$$\widehat{V}_{\hat{\beta}_{POLS}} = (X'X)^{-1}\left(\sum_{i=1}^{n} X_i'\hat{e}_i\hat{e}_i'X_i\right)(X'X)^{-1}$$

$$\sqrt{nT}(\hat{\beta}_{POLS} - \beta) \xrightarrow{d} N(0, V_\beta)$$

- Things are not clear cut if c_i can be correlated with x_{it}

Random effect: Treat c_i like another error term

- We assume that c_i is latent, but uncorrelated with x_i

Random effects assumptions

RE1 We assume $E[c_i|X_i] = 0$, $E[c_i] = 0$ and strict exogeneity $E[e_{it}|X_i, c_i] = 0$

RE2 rank $\left(E[X_i'\Omega^{-1}X_i]\right) = k$ (full column rank)

RE3 Conditionally spherical variance matrix: $E[e_i e_i'|X_i, c_i] = \sigma_e^2 I_T$ and $E[c_i^2|X_i] = \sigma_c^2$

- The covariance of e_{it} and v_{it} is

$$E[e_{it}e_{is}] = \begin{cases} 0 & t \neq s \\ \sigma_e^2 & t = s \end{cases}, \quad E[v_{it}v_{is}] = \begin{cases} \sigma_c^2 & t \neq s \\ \sigma_e^2 + \sigma_c^2 & t = s \end{cases}$$

- Composite error term v_{it} has a serial correlation! $\rightarrow$ GLS gives efficient estimates

Variance-covariance matrix of the v_{it}

- We can write

$$v_i = 1_T c_i + u_i$$

- Using this notation, we can define

$$\begin{aligned} E[v_i v_i'] &= E\left[(1_T c_i + e_i)(1_T c_i + e_i)'\right] = E[1_T 1_T' c_i^2 + e_i e_i'] \\ &= \sigma_c^2 1_T 1_T' + \sigma_e^2 I_T \end{aligned}$$

- $E[v_i v_i']$ has the following matrix representation

$$E[v_i v_i'] = \begin{pmatrix} \sigma_c^2 + \sigma_e^2 & \sigma_c^2 & \dots & \dots. \\ \sigma_c^2 & \sigma_c^2 + \sigma_e^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \sigma_c^2 & \sigma_c^2 + \sigma_e^2 \end{pmatrix} = \Omega$$

GLS estimator for RE

- Using Ω, we can write

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^{n} X_i'\widehat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{n} X_i'\widehat{\Omega}^{-1}y_i\right)$$

- Sample variance of $\left(\sum_{i=1}^{n} X_i'\widehat{\Omega}^{-1}X_i\right)^{-1}$.
- $\widehat{\Omega}$ can be obtained by estimating each component of the Ω matrix. σ_e^2 can be obtained by the unbiased estimate of $E[e_{it}^2]$ and σ_c^2 from that for $E[v_{it}v_{is}]$ for $t \neq s$.

$$\hat{\sigma}_e^2 = \frac{1}{nT-k}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{e}_{it}^2$$

$$\hat{\sigma}_c^2 = \frac{1}{nT(T-1)/2-k}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$$

RE GLS as a quasi-demeaned estimator

- Note that

$$E[v_i v_i'] = \sigma_c^2 1_T 1_T' + \sigma_e^2 I_T = T\sigma_c^2 1_T (\underbrace{1_T' 1_T}_{T})^{-1} 1_T' + \sigma_e^2 I_T = T\sigma_c^2 P_{1_T} + \sigma_e^2 I_T$$

- So the square root matrix of Ω, which I write P (s.t. $PP' = \Omega^{-1}$) is

$$P' = \frac{1}{\sigma_e}[I_T - \rho P_T] \quad \left(\rho = 1 - \sqrt{\frac{\sigma_e^2}{T\sigma_c^2 + \sigma_e^2}}\right)$$

- Thus, the transformed data generating process $\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{v}_{it}$, where $\tilde{y}_{it} = y_{it} - \rho\bar{y}_i$ has a spherical variances involving $\tilde{v}_{it}$ term.