Time-Series-Cross-Section Data

Jason Qiang Guo, New York

Why PCSE for TSCS Data?

Why PCSE for TSCS Data?

Basic Setting

- Assume we have a generic TCSC model

$$Y_{i,t} = \mathbf{X_{i,t}}\beta + \epsilon_{i,t};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T,$$

where $\mathbf{X_{i,t}}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).

Why PCSE for TSCS Data?

Basic Setting

- Assume we have a generic TCSC model

$$Y_{i,t} = \mathbf{X_{i,t}}\beta + \epsilon_{i,t};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T,$$

  where $\mathbf{X_{i,t}}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).

- Spherical errors for OLS to be optimal:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise} \end{cases}$$

Why PCSE for TSCS Data?

Basic Setting

- Assume we have a generic TCSC model

$$Y_{i,t} = \mathbf{X_{i,t}}\beta + \epsilon_{i,t};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T,$$

  where $\mathbf{X_{i,t}}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).

- Spherical errors for OLS to be optimal:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise} \end{cases}$$

- Question: what does this assumption of spherical assumption imply about error processes?

Why PCSE for TSCS Data?

Basic Setting

- Assume we have a generic TCSC model

$$Y_{i,t} = \mathbf{X_{i,t}}\beta + \epsilon_{i,t};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T,$$

  where $\mathbf{X_{i,t}}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).

- Spherical errors for OLS to be optimal:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise} \end{cases}$$

- Question: what does this assumption of spherical assumption imply about error processes?
  - All the error processes have the same variance (homoscedasticity)
  - Errors for one unit are unrelated to errors for every other unit (no spatial correlation)
  - Errors for a particular unit at one time are unrelated to errors for that unit at all other times (no serial correlation)

Why PCSE for TSCS Data?

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$
- Given that error processes of TCSC data are unlikely spherical, OLS estimator is not efficient anymore (and all the tests are also invalid) as the variance-covariance matrix is Ω instead of $\sigma^2 \otimes I_{NT \times NT}$

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- Given that error processes of TCSC data are unlikely spherical, OLS estimator is not efficient anymore (and all the tests are also invalid) as the variance-covariance matrix is Ω instead of $\sigma^2 \otimes I_{NT\times NT}$
- Now the vcov matrix for the OLS estimator becomes $(X'X)^{-1}X'\Omega X(X'X)^{-1}$

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- Given that error processes of TCSC data are unlikely spherical, OLS estimator is not efficient anymore (and all the tests are also invalid) as the variance-covariance matrix is Ω instead of $\sigma^2 \otimes I_{NT\times NT}$
- Now the vcov matrix for the OLS estimator becomes $(X'X)^{-1}X'\Omega X(X'X)^{-1}$
- Suppose there is an estimator $\hat{\Omega}$ for Ω, then we get an FGLS estimator.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$
- Given that error processes of TCSC data are unlikely spherical, OLS estimator is not efficient anymore (and all the tests are also invalid) as the variance-covariance matrix is Ω instead of $\sigma^2 \otimes I_{NT \times NT}$
- Now the vcov matrix for the OLS estimator becomes $(X'X)^{-1} X' \Omega X (X'X)^{-1}$
- Suppose there is an estimator $\hat{\Omega}$ for Ω, then we get an FGLS estimator.
- If the structure of error processes is known, suppose we have $C \Omega C' = \sigma^2 \otimes I_{NT \times NT}$, we can transform the model to $Y^* = X^* \beta + \epsilon^*$, where $\epsilon^* = C\epsilon$, $Y^* = CY$ and $X^* = CX$ (Weighted Least Square Approach!), then we get BLUE $\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} Y^*$ under heteroscedasticity/serial correlation by construction.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Recall the variance-covariance matrix for the OLS estimator is $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- Given that error processes of TCSC data are unlikely spherical, OLS estimator is not efficient anymore (and all the tests are also invalid) as the variance-covariance matrix is Ω instead of $\sigma^2 \otimes I_{NT\times NT}$
- Now the vcov matrix for the OLS estimator becomes $(X'X)^{-1}X'\Omega X(X'X)^{-1}$
- Suppose there is an estimator $\hat{\Omega}$ for Ω, then we get an FGLS estimator.
- If the structure of error processes is known, suppose we have $C\Omega C' = \sigma^2 \otimes I_{NT\times NT}$, we can transform the model to $Y^* = X^*\beta + \epsilon^*$, where $\epsilon^* = C\epsilon$, $Y^* = CY$ and $X^* = CX$ (Weighted Least Square Approach!), then we get BLUE $\hat{\beta} = ({X^*}'X^*)^{-1}{X^*}'Y^*$ under heteroscedasticity/serial correlation by construction.
- The problem is that we usually don't know the structure of the error processes, therefore finding an estimator for Ω is critical.

Why PCSE for TSCS Data?

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Parks' (1967) correction for contemporaneously correlated errors:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2_{i,j} & \text{and } s = t \\ 0 & \text{otherwise} \end{cases}$$

which requires the estimation of $N(N+1)/2$ parameters using NT observations.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Parks' (1967) correction for contemporaneously correlated errors:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2_{i,j} & \text{and } s = t \\ 0 & \text{otherwise} \end{cases}$$

  which requires the estimation of $N(N+1)/2$ parameters using NT observations.

- This means if T is not significantly larger than N, proper estimation of Ω would be impossible and the finite sample properties of FGLS estimator therefore is unknown. The efficiency gain of FGLS is unclear.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Parks' (1967) correction for contemporaneously correlated errors:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2_{i,j} & \text{and } s = t \\ 0 & \text{otherwise} \end{cases}$$

  which requires the estimation of $N(N+1)/2$ parameters using NT observations.

- This means if T is not significantly larger than N, proper estimation of Ω would be impossible and the finite sample properties of FGLS estimator therefore is unknown. The efficiency gain of FGLS is unclear.

- Park's correction for serial correlation: assume $\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ is *i.i.d.*.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Parks' (1967) correction for contemporaneously correlated errors:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2_{i,j} & \text{and } s = t \\ 0 & \text{otherwise} \end{cases}$$

  which requires the estimation of $N(N+1)/2$ parameters using NT observations.

- This means if T is not significantly larger than N, proper estimation of Ω would be impossible and the finite sample properties of FGLS estimator therefore is unknown. The efficiency gain of FGLS is unclear.
- Park's correction for serial correlation: assume $\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ is *i.i.d.*.
- The FGLS correction for unit-specific serially correlated errors, used by Parks, is likely to cause more serious underestimates of variability. The essence of the problem is that each ρ_i, is estimated using an autoregression based on only T observations and such estimates are biased downward.

Why PCSE for TSCS Data?

Methods before PCSE was invented: Feasible Generalized Least Squares

- Parks' (1967) correction for contemporaneously correlated errors:

$$E(\epsilon_{i,t}, \epsilon_{j,s}) = \begin{cases} \sigma^2_{i,j} & \text{and } s = t \\ 0 & \text{otherwise} \end{cases}$$

  which requires the estimation of $N(N+1)/2$ parameters using NT observations.
- This means if T is not significantly larger than N, proper estimation of Ω would be impossible and the finite sample properties of FGLS estimator therefore is unknown. The efficiency gain of FGLS is unclear.
- Park's correction for serial correlation: assume $\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{i,t}$, where $\nu_{i,t}$ is *i.i.d.*.
- The FGLS correction for unit-specific serially correlated errors, used by Parks, is likely to cause more serious underestimates of variability. The essence of the problem is that each ρ_i, is estimated using an autoregression based on only T observations and such estimates are biased downward.
- There is also a substantive concern for the use of FGLS correction for unit-specific serial correlation. If we assume β do not vary across units, then why does not ρ show similar pooling?

Why PCSE for TSCS Data?

PCSE (Beck and Katz (1995), the most cited paper in political methodology and the second most cited paper in political science!)

Why PCSE for TSCS Data?

PCSE (Beck and Katz (1995), the most cited paper in political methodology and the second most cited paper in political science!)

- For panel data with contemporaneously correlated and panel heteroscedastitic errors, Ω is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\sum$, along the diagonal.

Why PCSE for TSCS Data?

PCSE (Beck and Katz (1995), the most cited paper in political methodology and the second most cited paper in political science!)

- For panel data with contemporaneously correlated and panel heteroscedastitic errors, Ω is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\sum$, along the diagonal.
- Given the consistency of OLS estimator, we can use OLS residuals from that estimation to provide a consistent estimator of $\sum$. We can estimate a typical element of $\sum$ by

$$\hat{\sum}_{i,j} = \frac{\sum_{t=1}^{T_{i,j}} e_{i,t} e_{j,t}}{T_{i,j}},$$

  where $e_{i,t}$ is the OLS residual for unit i at time t, and $T_{i,j} = T \ \ \forall i = 1, \ldots, N$ for balanced data. And we can simplify this to $\hat{\sum} = \frac{\mathbf{E}'\mathbf{E}}{T}$, where $\mathbf{E}$ is the $T \times N$ residual matrix.

Why PCSE for TSCS Data?

PCSE (Beck and Katz (1995), the most cited paper in political methodology and the second most cited paper in political science!)

- For panel data with contemporaneously correlated and panel heteroscedastitic errors, Ω is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\sum$, along the diagonal.
- Given the consistency of OLS estimator, we can use OLS residuals from that estimation to provide a consistent estimator of $\sum$. We can estimate a typical element of $\sum$ by
$$\hat{\sum}_{i,j} = \frac{\sum_{t=1}^{T_{i,j}} e_{i,t} e_{j,t}}{T_{i,j}},$$
where $e_{i,t}$ is the OLS residual for unit i at time t, and $T_{i,j} = T \ \ \forall i = 1, \ldots, N$ for balanced data. And we can simplify this to $\hat{\sum} = \frac{\mathbf{E}'\mathbf{E}}{T}$, where $\mathbf{E}$ is the $T \times N$ residual matrix.
- And therefore $PCSE = (X'X)^{-1}X'(\frac{\mathbf{E}'\mathbf{E}}{\mathbf{T}} \otimes I_T)X(X'X)^{-1}$

Why PCSE for TSCS Data?

PCSE (Beck and Katz (1995), the most cited paper in political methodology and the second most cited paper in political science!)

- For panel data with contemporaneously correlated and panel heteroscedastitic errors, Ω is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, $\sum$, along the diagonal.
- Given the consistency of OLS estimator, we can use OLS residuals from that estimation to provide a consistent estimator of $\sum$. We can estimate a typical element of $\sum$ by

$$\hat{\sum}_{i,j} = \frac{\sum_{t=1}^{T_{i,j}} e_{i,t} e_{j,t}}{T_{i,j}},$$

where $e_{i,t}$ is the OLS residual for unit i at time t, and $T_{i,j} = T \ \ \forall i = 1, \ldots, N$ for balanced data. And we can simplify this to $\hat{\sum} = \frac{\mathbf{E}'\mathbf{E}}{T}$, where $\mathbf{E}$ is the $T \times N$ residual matrix.

- And therefore $PCSE = (X'X)^{-1} X' (\frac{\mathbf{E}'\mathbf{E}}{\mathbf{T}} \otimes I_T) X (X'X)^{-1}$
- What if we have autocorrelation in the data? Eliminate autocorrelation through model specification change (manipulate lag term) or Prais-Winston transformation, estimate the model and then apply PCSE