Modeling of Time-Series Cross-Sectional / Panel Data

David Carlson

May 9, 2022

TSCS and Panel Data

- Repeated measurements over time, across units of observation

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel
- It is very important to note which of the modeling strategies (most of them) assume balance; this is often overlooked

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel
- It is very important to note which of the modeling strategies (most of them) assume balance; this is often overlooked
- We have to consider all of the issues discussed in temporal, and add to that heterogeneity at the unit level

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel
- It is very important to note which of the modeling strategies (most of them) assume balance; this is often overlooked
- We have to consider all of the issues discussed in temporal, and add to that heterogeneity at the unit level
- This heterogeneity could be different baselines, different slopes (treatment effects), etc.

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel
- It is very important to note which of the modeling strategies (most of them) assume balance; this is often overlooked
- We have to consider all of the issues discussed in temporal, and add to that heterogeneity at the unit level
- This heterogeneity could be different baselines, different slopes (treatment effects), etc.
- We also need to consider auto-correlation, contagion, etc. across units as well as over time

TSCS and Panel Data

- Repeated measurements over time, across units of observation
- There are no agreed upon defining characteristics between TSCS and panel data, but for our purposes we will refer to large N, mostly balanced data as panel
- It is very important to note which of the modeling strategies (most of them) assume balance; this is often overlooked
- We have to consider all of the issues discussed in temporal, and add to that heterogeneity at the unit level
- This heterogeneity could be different baselines, different slopes (treatment effects), etc.
- We also need to consider auto-correlation, contagion, etc. across units as well as over time
- Unfortunately, there is no one best solution; you need to carefully consider the level of aggregation, the likely DGP, etc.

One- and Two-Way Fixed Effects Models

- A ‘fixed effect’ is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed

One- and Two-Way Fixed Effects Models

- A ‘fixed effect’ is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed
- In panel data, we can include a ‘wave’ (temporal) effect and/or a unit effect

One- and Two-Way Fixed Effects Models

- A ‘fixed effect’ is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed
- In panel data, we can include a ‘wave’ (temporal) effect and/or a unit effect
- Most common to use two-way (both temporal and unit) for panel data

One- and Two-Way Fixed Effects Models

- A ‘fixed effect’ is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed
- In panel data, we can include a ‘wave’ (temporal) effect and/or a unit effect
- Most common to use two-way (both temporal and unit) for panel data
- For TSCS, two-way is also common, but the interpretability and framework are hurt by multiple observations for a unit-time, or lack of balance

One- and Two-Way Fixed Effects Models

- A 'fixed effect' is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed
- In panel data, we can include a 'wave' (temporal) effect and/or a unit effect
- Most common to use two-way (both temporal and unit) for panel data
- For TSCS, two-way is also common, but the interpretability and framework are hurt by multiple observations for a unit-time, or lack of balance
- Fixed effects control for heterogeneity in the data

One- and Two-Way Fixed Effects Models

- A ‘fixed effect’ is colloquially simply an indicator, a dummy variable — but really they are all effects that are not assumed randomly distributed
- In panel data, we can include a ‘wave’ (temporal) effect and/or a unit effect
- Most common to use two-way (both temporal and unit) for panel data
- For TSCS, two-way is also common, but the interpretability and framework are hurt by multiple observations for a unit-time, or lack of balance
- Fixed effects control for heterogeneity in the data
- The inclusion of unit and time fixed effects accounts for both unit-specific (but time-invariant) and time-specific (but unit-invariant) unobserved confounders in a flexible manner

FE (cont.)

- Their statistical power can be very low with small N, because of the loss in degrees of freedom

FE (cont.)

- Their statistical power can be very low with small N, because of the loss in degrees of freedom
- Unlike random/mixed effects, they do not have distributional assumptions

FE (cont.)

- Their statistical power can be very low with small N, because of the loss in degrees of freedom
- Unlike random/mixed effects, they do not have distributional assumptions
- These two points mean that there is no information 'borrowed' from other unit-times, and the variation that would be explained by your independent variable is largely absorbed

FE (cont.)

- Their statistical power can be very low with small N, because of the loss in degrees of freedom
- Unlike random/mixed effects, they do not have distributional assumptions
- These two points mean that there is no information 'borrowed' from other unit-times, and the variation that would be explained by your independent variable is largely absorbed
- These have great asymptotic properties, but a lot of research on the actual practice show they perform rather poorly, even when assumptions are met

FE (cont.)

- Their statistical power can be very low with small N, because of the loss in degrees of freedom
- Unlike random/mixed effects, they do not have distributional assumptions
- These two points mean that there is no information 'borrowed' from other unit-times, and the variation that would be explained by your independent variable is largely absorbed
- These have great asymptotic properties, but a lot of research on the actual practice show they perform rather poorly, even when assumptions are met
- Very simple to run; just a linear model with factors

FE Assumptions

- Linearity (although one could nonparametrically adjust for unit-specific (time-specific) unobserved confounders by matching a treated observation with control observations of the same unit (time period), no other observation shares the same unit and time indices. Thus, the 2FE estimator critically relies upon the linearity assumption for its simultaneous adjustment for the two types of unobserved confounders)

FE Assumptions

- Linearity (although one could nonparametrically adjust for unit-specific (time-specific) unobserved confounders by matching a treated observation with control observations of the same unit (time period), no other observation shares the same unit and time indices. Thus, the 2FE estimator critically relies upon the linearity assumption for its simultaneous adjustment for the two types of unobserved confounders)
- Separability (precisely when their respective convex hulls are disjoint (colloquially, do not overlap); means that the marginal rate of substitution between any pair of primary inputs is independent of the amount of intermediate inputs used)

FE Assumptions

- Linearity (although one could nonparametrically adjust for unit-specific (time-specific) unobserved confounders by matching a treated observation with control observations of the same unit (time period), no other observation shares the same unit and time indices. Thus, the 2FE estimator critically relies upon the linearity assumption for its simultaneous adjustment for the two types of unobserved confounders)
- Separability (precisely when their respective convex hulls are disjoint (colloquially, do not overlap); means that the marginal rate of substitution between any pair of primary inputs is independent of the amount of intermediate inputs used)
- Additivity (means that the effect of one independent variable(s) on the dependent variable does NOT depend on the value of another independent variable(s))

FE Assumptions

- Linearity (although one could nonparametrically adjust for unit-specific (time-specific) unobserved confounders by matching a treated observation with control observations of the same unit (time period), no other observation shares the same unit and time indices. Thus, the 2FE estimator critically relies upon the linearity assumption for its simultaneous adjustment for the two types of unobserved confounders)
- Separability (precisely when their respective convex hulls are disjoint (colloquially, do not overlap); means that the marginal rate of substitution between any pair of primary inputs is independent of the amount of intermediate inputs used)
- Additivity (means that the effect of one independent variable(s) on the dependent variable does NOT depend on the value of another independent variable(s))
- Functional form assumption (are we modeling the DGP?)

FE Assumptions

- Linearity (although one could nonparametrically adjust for unit-specific (time-specific) unobserved confounders by matching a treated observation with control observations of the same unit (time period), no other observation shares the same unit and time indices. Thus, the 2FE estimator critically relies upon the linearity assumption for its simultaneous adjustment for the two types of unobserved confounders)
- Separability (precisely when their respective convex hulls are disjoint (colloquially, do not overlap); means that the marginal rate of substitution between any pair of primary inputs is independent of the amount of intermediate inputs used)
- Additivity (means that the effect of one independent variable(s) on the dependent variable does NOT depend on the value of another independent variable(s))
- Functional form assumption (are we modeling the DGP?)
- Adjustment for the two types of unobserved confounders cannot be done nonparametrically under the 2FE framework

Difference-in-Differences Estimator

- DiD estimators have traditionally been used to analyze the effect of a treatment at different times

Difference-in-Differences Estimator

- DiD estimators have traditionally been used to analyze the effect of a treatment at different times
- There is absolutely no reason not to use them for continuous treatments though

Difference-in-Differences Estimator

- DiD estimators have traditionally been used to analyze the effect of a treatment at different times
- There is absolutely no reason not to use them for continuous treatments though
- Basically, we interact temporal fixed effects (but not the pre-treatment, or baseline year(s)) with the variable of interest (and include the constituent terms of time), and include unit fixed effects

Difference-in-Differences Estimator

- DiD estimators have traditionally been used to analyze the effect of a treatment at different times
- There is absolutely no reason not to use them for continuous treatments though
- Basically, we interact temporal fixed effects (but not the pre-treatment, or baseline year(s)) with the variable of interest (and include the constituent terms of time), and include unit fixed effects
- This is an ideal model for for causal identification when treatment occured after the beginning of your data

Difference-in-Differences Estimator

- DiD estimators have traditionally been used to analyze the effect of a treatment at different times
- There is absolutely no reason not to use them for continuous treatments though
- Basically, we interact temporal fixed effects (but not the pre-treatment, or baseline year(s)) with the variable of interest (and include the constituent terms of time), and include unit fixed effects
- This is an ideal model for for causal identification when treatment occured after the beginning of your data
- However, we need to assume parallel trends, i.e. the units (municipalities) have parallel trends in the outcome if it were not for treatment

Adjusting Standard Errors

- This is very, very common, and you will likely be asked to do it by reviewers, committee members, etc.

Adjusting Standard Errors

- This is very, very common, and you will likely be asked to do it by reviewers, committee members, etc.
- There is a ton of work arguing that these are misleading adjustments

Adjusting Standard Errors

- This is very, very common, and you will likely be asked to do it by reviewers, committee members, etc.
- There is a ton of work arguing that these are misleading adjustments
- They do not change the point estimates, only the SEs

Adjusting Standard Errors

- This is very, very common, and you will likely be asked to do it by reviewers, committee members, etc.
- There is a ton of work arguing that these are misleading adjustments
- They do not change the point estimates, only the SEs
- If you need to include them, include non-adjusted results (in an appendix at least)

Adjusting Standard Errors

- This is very, very common, and you will likely be asked to do it by reviewers, committee members, etc.
- There is a ton of work arguing that these are misleading adjustments
- They do not change the point estimates, only the SEs
- If you need to include them, include non-adjusted results (in an appendix at least)
- When corrected SEs and un-corrected disagree, you very likely have a mis-specified model

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)
- The latter takes time into effect as a measure of distance, the former is agnostic about the effects but still corrects for each 'cluster'

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)
- The latter takes time into effect as a measure of distance, the former is agnostic about the effects but still corrects for each ‘cluster’
- RCSE are extremely popular, and you will be asked for revisions that include them

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)
- The latter takes time into effect as a measure of distance, the former is agnostic about the effects but still corrects for each 'cluster'
- RCSE are extremely popular, and you will be asked for revisions that include them
- Generally speaking, though, they suffer the same problems as robust standard errors

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)
- The latter takes time into effect as a measure of distance, the former is agnostic about the effects but still corrects for each ‘cluster’
- RCSE are extremely popular, and you will be asked for revisions that include them
- Generally speaking, though, they suffer the same problems as robust standard errors
- Furthermore, it is WIDELY assumed that they will increase SEs, but this is very often not true in practice

Adjusting SEs (cont.)

- There are two main adjustments; robust clustered standard errors (RCSE) and panel corrected standard errors (PCSE)
- The latter takes time into effect as a measure of distance, the former is agnostic about the effects but still corrects for each ‘cluster’
- RCSE are extremely popular, and you will be asked for revisions that include them
- Generally speaking, though, they suffer the same problems as robust standard errors
- Furthermore, it is WIDELY assumed that they will increase SEs, but this is very often not true in practice
- Easy in `R`, but for the above reasons not as simple as a flag as in STATA

Lagged Dependent Variable Models

- On the right-hand side of the model, we can include a lagged DV

Lagged Dependent Variable Models

- On the right-hand side of the model, we can include a lagged DV
- The lagged DV corrects for auto-correlation and is often the most predictive variable

Lagged Dependent Variable Models

- On the right-hand side of the model, we can include a lagged DV
- The lagged DV corrects for auto-correlation and is often the most predictive variable
- However, this makes interpretation of IV estimate difficult

Lagged Dependent Variable Models

- On the right-hand side of the model, we can include a lagged DV
- The lagged DV corrects for auto-correlation and is often the most predictive variable
- However, this makes interpretation of IV estimate difficult
- This also absorbs variation that may be meaningful

Lagged Dependent Variable Models

- On the right-hand side of the model, we can include a lagged DV
- The lagged DV corrects for auto-correlation and is often the most predictive variable
- However, this makes interpretation of IV estimate difficult
- This also absorbs variation that may be meaningful
- Great for prediction, but not for inference; nevertheless this is another robustness check that is often asked for

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$
- This will often (but not always) make a non-stationary series stationary

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$
- This will often (but not always) make a non-stationary series stationary
- The interpretation of β is now differences on differences

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$
- This will often (but not always) make a non-stationary series stationary
- The interpretation of β is now differences on differences
- We are implicitly controlling for baseline heterogeneity in the outcomes across units

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$
- This will often (but not always) make a non-stationary series stationary
- The interpretation of β is now differences on differences
- We are implicitly controlling for baseline heterogeneity in the outcomes across units
- Model captures the average treatment effect (ATE)

First-Differencing

- Rather than model y as a function of x at time t, we can model $y_t - y_{t-1}$ on $x_t - x_{t-1}$
- This will often (but not always) make a non-stationary series stationary
- The interpretation of β is now differences on differences
- We are implicitly controlling for baseline heterogeneity in the outcomes across units
- Model captures the average treatment effect (ATE)
- First-difference estimators are designed to control for unobserved covariates in panel data, so we require less assumptions about the underlying mean function

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things
  - We can borrow information from other units to find a generalized pattern across units

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things
  - We can borrow information from other units to find a generalized pattern across units
  - The statistical power is much higher

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things
  - We can borrow information from other units to find a generalized pattern across units
  - The statistical power is much higher
  - The power means we can get false positives, but generally only under mis-specification

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things
  - We can borrow information from other units to find a generalized pattern across units
  - The statistical power is much higher
  - The power means we can get false positives, but generally only under mis-specification
  - We can directly estimate the variation of the underlying distribution, meaning we can inspect which levels need variation accounted for

Mixed (Random and Fixed) Effects

- We can use these in much the same way (and a lot more) as FEs
- Random effects have a distributional assumption
- This distribution buys us a lot of things
  - We can borrow information from other units to find a generalized pattern across units
  - The statistical power is much higher
  - The power means we can get false positives, but generally only under mis-specification
  - We can directly estimate the variation of the underlying distribution, meaning we can inspect which levels need variation accounted for
  - We no longer make the assumption of separability, and estimate the covariation of the parameters

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation
- You can model a small number of units

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation
- You can model a small number of units
- You can model a small number of observations

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation
- You can model a small number of units
- You can model a small number of observations
- You can add predictors to the random effects for more precision (contextual effects)

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation
- You can model a small number of units
- You can model a small number of observations
- You can add predictors to the random effects for more precision (contextual effects)
- You can also model any slope coefficient randomly, allowing it to vary at any desired level

Mixed (cont.)

- The distribution, in general, assumes that the population under study is representative of the group-of-interest, in that the effects are random realizations of the randomly distributed population (and temporal) heterogeneity (prevalancy kind of)
- We will generally want to go Bayesian, for a number of reasons we will cover later, but basically it allows for much more control over the model and a much more flexible DGP
- Random effects do not require balance
- You can model units that only have one observation
- You can model a small number of units
- You can model a small number of observations
- You can add predictors to the random effects for more precision (contextual effects)
- You can also model any slope coefficient randomly, allowing it to vary at any desired level
- Assumes that errors are uncorrelated with regressors