TSCS Analysis

Binary Outcome TSCS Models

Jude C. Hays
jch61@pitt.edu

April 14, 2021

Spatiotemporal Multivariate Qualitative Models

Important questions: (observed) recursive? (latent) simultaneous? mixed process?

A general form for a pair of binary outcomes (here as probit) is given as:

$$
\begin{aligned}
y_1^* &= \beta_0 + \beta_1 x_1 + \beta_2 y_2 + \beta_3 y_2^* + u \\
y_2^* &= \gamma_0 + \gamma_1 x_1 + \gamma_2 y_1 + \gamma_3 y_1^* + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma)
\end{aligned}
$$

Spatiotemporal Multivariate Qualitative Models

Important questions: (observed) recursive? (latent) simultaneous? mixed process?

A general form for a pair of binary outcomes (here as probit) is given as:

$$
\begin{aligned}
y_1^* &= \beta_0 + \beta_1 x_1 + \beta_2 y_2 + \beta_3 y_2^* + u \\
y_2^* &= \gamma_0 + \gamma_1 x_1 + \gamma_2 y_1 + \gamma_3 y_1^* + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma)
\end{aligned}
$$

This nests several models:

1. If $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits

Spatiotemporal Multivariate Qualitative Models

Important questions: (observed) recursive? (latent) simultaneous? mixed process?

A general form for a pair of binary outcomes (here as probit) is given as:

$$y_1^* = \beta_0 + \beta_1 x_1 + \beta_2 y_2 + \beta_3 y_2^* + u$$
$$y_2^* = \gamma_0 + \gamma_1 x_1 + \gamma_2 y_1 + \gamma_3 y_1^* + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma)$$

This nests several models:

1. If $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits
2. If γ_2 or $\beta_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits (recursive)

Spatiotemporal Multivariate Qualitative Models

Important questions: (observed) recursive? (latent) simultaneous? mixed process?

A general form for a pair of binary outcomes (here as probit) is given as:

$$y_1^* = \beta_0 + \beta_1 x_1 + \beta_2 y_2 + \beta_3 y_2^* + u$$
$$y_2^* = \gamma_0 + \gamma_1 x_1 + \gamma_2 y_1 + \gamma_3 y_1^* + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma)$$

This nests several models:

1. If $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits
2. If γ_2 or $\beta_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits (recursive)
3. If γ_2 or $\beta_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} \neq 0$ bi-probit

Spatiotemporal Multivariate Qualitative Models

Important questions: (observed) recursive? (latent) simultaneous? mixed process?

A general form for a pair of binary outcomes (here as probit) is given as:

$$y_1^* = \beta_0 + \beta_1 x_1 + \beta_2 y_2 + \beta_3 y_2^* + u$$
$$y_2^* = \gamma_0 + \gamma_1 x_1 + \gamma_2 y_1 + \gamma_3 y_1^* + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma)$$

This nests several models:

1. If $\beta_2 = \gamma_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits
2. If γ_2 or $\beta_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} = 0$ ind. probits (recursive)
3. If γ_2 or $\beta_2 = \beta_3 = \gamma_3 = 0$ and $\rho_{uv} \neq 0$ bi-probit
4. If $\beta_3 \neq 0$ and $\gamma_3 \neq 0$ simultaneous equation model

Spatiotemporal Multivariate Qualitative Models

When both latent and realized qualitative outcome are in the model, additional restrictions may be needed for logical consistency.

Spatiotemporal Multivariate Qualitative Models

When both latent and realized qualitative outcome are in the model, additional restrictions may be needed for logical consistency.Consider

$$y^* = \beta_1 x + \beta_2 y + u,$$

is logically inconsistent unless $\beta_2 = 0$. To see, when $y = 1$ we have $\beta_1 x + \beta_2 < u$ and when $y = 0$ then $\beta_1 x \geq u$, where the probabilities only sum to 1 when $\beta = 0$.

Spatiotemporal Multivariate Qualitative Models

When both latent and realized qualitative outcome are in the model, additional restrictions may be needed for logical consistency.Consider

$$y^* = \beta_1 x + \beta_2 y + u,$$

is logically inconsistent unless $\beta_2 = 0$. To see, when $y = 1$ we have $\beta_1 x + \beta_2 < u$ and when $y = 0$ then $\beta_1 x \geq u$, where the probabilities only sum to 1 when $\beta = 0$.

More general than it might seem, consider

$$\begin{aligned} y_1 &= \beta_1 y_2 + \beta_2 x_1 + u \\ y_2^* &= \gamma_1 y_1 + \gamma_2 x_2 + v = \gamma_1 \beta_1 y_2 + \gamma_1 \beta_2 x_1 + \gamma_2 x_2 + \gamma_1 u + v \end{aligned}$$

Spatiotemporal Multivariate Qualitative Models

When both latent and realized qualitative outcome are in the model, additional restrictions may be needed for logical consistency.Consider

$$y^* = \beta_1 x + \beta_2 y + u,$$

is logically inconsistent unless $\beta_2 = 0$. To see, when $y = 1$ we have $\beta_1 x + \beta_2 < u$ and when $y = 0$ then $\beta_1 x \geq u$, where the probabilities only sum to 1 when $\beta = 0$.

More general than it might seem, consider

$$\begin{aligned} y_1 &= \beta_1 y_2 + \beta_2 x_1 + u \\ y_2^* &= \gamma_1 y_1 + \gamma_2 x_2 + v = \gamma_1 \beta_1 y_2 + \gamma_1 \beta_2 x_1 + \gamma_2 x_2 + \gamma_1 u + v \end{aligned}$$

Only (logically) consistent if $\gamma_1 \beta_1 = 0$

If, however, restrictions are made to ensure that the model is recursive:

$$y_1^* = \beta_0 + \beta_1 x_1 + \beta_2 y_2 + u$$
$$y_2^* = \gamma_0 + \gamma_1 x_1 + v, \text{ where } (u, v) \sim \mathcal{N}(0, \Sigma),$$

then estimation proceeds without issue any additional fixes.

Spatiotemporal Multivariate Qualitative Models

Thus, for simplicity, let's assume $\beta_2 = 0$ and just focus on the correlated errors. To estimate such a model we need to denote a joint distribution of (u, v) – say $F(\cdot, \cdot)$ – and we can define joint probabilities for the outcome profiles:

$$P_{11} = \Pr(y_1 = 1, y_2 = 1) = F[(\beta_1 x_1, \beta_2 x_2); \rho]$$
$$P_{10} = \Pr(y_1 = 1, y_2 = 0) = F[(\beta_1 x_1, -\beta_2 x_2); \rho]$$
$$P_{01} = \Pr(y_1 = 0, y_2 = 1) = F[(-\beta_1 x_1, \beta_2 x_2); \rho]$$
$$P_{00} = \Pr(y_1 = 0, y_2 = 0) = F[(-\beta_1 x_1, -\beta_2 x_2); \rho]$$

Which we could specify a likelihood function for as:

$$L(\beta_1, \beta_2) = \prod P_{11}^{y_1 y_2} P_{11}^{y_1(1-y_2)} P_{11}^{(1-y_1)y_2} P_{11}^{(1-y_1)(1-y_2)}$$

Spatiotemporal Multivariate Qualitative Models

To back up for a second, maximizing the likelihood from the last slide involves the evaluation of double integrals (the bivariate normal CDF). For example

$$P_{00} = \int_{-\infty}^{-\beta_1 x_1} \int_{-\infty}^{-\beta_2 x_2} \phi((u,v)'; \Sigma) du dv = \Phi(-\beta_1 x_1, -\beta_2 x_2; \rho),$$

or

$$P_{11} = \int_{-\beta_1 x_1}^{\infty} \int_{-\beta_2 x_2}^{\infty} \phi((u,v)'; \Sigma) du dv = \Phi(\beta_1 x_1, \beta_2 x_2; \rho),$$

Can generalize using $q_j = 2y_{ij} - 1$

$$\ell = \sum \log \Phi(q_j \beta_1 x_1, q_j \beta_2 x_2; q_1 q_2 \rho),$$

Space-Time Dependence via Latents

This represents a “pure” strategy to model autoregressive dependence in space and time (Franseze, Hays, Cook 2016). However. . .

Space-Time Dependence via Latents

This represents a "pure" strategy to model autoregressive dependence in space and time (Franseze, Hays, Cook 2016). However. . .

- In terms of specification multivariate models just extend the number of equations, but for estimation this becomes a "ferocious maximization problem"

Space-Time Dependence via Latents

This represents a “pure” strategy to model autoregressive dependence in space and time (Franseze, Hays, Cook 2016). However…

- In terms of specification multivariate models just extend the number of equations, but for estimation this becomes a “ferocious maximization problem”

Space-Time Dependence via Latents

This represents a "pure" strategy to model autoregressive dependence in space and time (Franseze, Hays, Cook 2016). However...

- In terms of specification multivariate models just extend the number of equations, but for estimation this becomes a "ferocious maximization problem"

Therefore we discuss some simple strategies (Beck, Katz, Tucker 1998; Carter and Signorino 2010)

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

- If interested in the $\text{pr}(y = 1)$ then logit, probit, etc.

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

- If interested in the $\text{pr}(y = 1)$ then logit, probit, etc.

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

- If interested in the pr$(y = 1)$ then logit, probit, etc.
- If interested in the length (of the spell) of 0's or 1's then survival/event history analysis

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

- If interested in the pr$(y = 1)$ then logit, probit, etc.
- If interested in the length (of the spell) of 0's or 1's then survival/event history analysis

On time: BTSCS is duration data

Underlying data structure:

t	1	2	3	4	5	6	7
y	0	0	1	1	1	1	0

Looks familiar?

- If interested in the pr$(y = 1)$ then logit, probit, etc.
- If interested in the length (of the spell) of 0's or 1's then survival/event history analysis

Therefore we can use strategies from duration models to account for time dependence in repeated-measure binary outcome data

Discrete-Time Hazard Models

- Discrete-time hazard models are used when the time to an event is either inherently discrete or group into discrete intervals of time ("interval censoring").
- With interval censoring, we model the probability that an event occurs within a particular interval with logit or probit models.
- The baseline hazard (i.e., the hazard function when all covariates are set to zero) can be modeled using either time-period dummies, splines, or polynomials.
- Polynomials in time are the easiest method to implement and interpret.
  - Time-period dummies are the most flexible, but they eat up degrees of freedom and commonly suffer from separation problems.
  - Splines are more difficult to implement correctly and rarely interpreted.
- Typically a cubic polynomial in time is sufficient to capture the baseline hazard.

Underlying maths: failure distribution

Define the CDF as

$$F(t) = \int_0^t f(u)d(u) = \text{pr}(T \leq t),$$

which is the probability that survival time T is less than or equal to t (sometimes called the failure distribution).

Underlying maths: failure distribution

Define the CDF as

$$F(t) = \int_0^t f(u)d(u) = \text{pr}(T \leq t),$$

which is the probability that survival time T is less than or equal to t (sometimes called the failure distribution).The PDF is

$$f(t) = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{\text{pr}(t \leq T \leq (t + \Delta t))}{\Delta t},$$

which is the conditional (or instantaneous) failure rate, that is, the probability that failure (an event) occurs within an infinitesimally small time.

Underlying maths: survivor function

The complement to the failure distribution is the, wait for it, survivor function:

$$S(t) = 1 - F(t) = \mathsf{pr}(T \geq t),$$

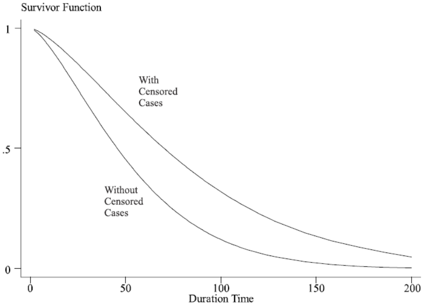
which gives us the – strictly non-increasing – proportion of survivors at time t

Underlying maths: survivor function

The complement to the failure distribution is the, wait for it, survivor function:

$$S(t) = 1 - F(t) = \text{pr}(T \geq t),$$

which gives us the – strictly non-increasing – proportion of survivors at time t



Underlying maths: hazard rate

Now we have the parts necessary to define the risk of failure, that is, the hazard rate

$$h(t) = \frac{f(t)}{S(t)} = \frac{\text{pr}(t \leq T \leq (t + \Delta t)|T \leq t)}{\Delta t}$$

which is the conditional failure rate – e.g., Given that the US has been a democracy since 1776, what are the chances it will transition to autocracy in 2017?

Underlying maths: hazard rate

Now we have the parts necessary to define the risk of failure, that is, the hazard rate

$$h(t) = \frac{f(t)}{S(t)} = \frac{\text{pr}(t \leq T \leq (t + \Delta t)|T \leq t)}{\Delta t}$$

which is the conditional failure rate – e.g., Given that the US has been a democracy since 1776, what are the chances it will transition to autocracy in 2017?

Discrete-Time Hazard Models

- The baseline hazard (i.e., the hazard function when all covariates are set to zero) can be modeled using either time-period dummies, splines, or polynomials.
- Polynomials in time are the easiest method to implement and interpret.
  - time-since-event counter, time-since-event counter^2, time-since-event counter^3
- Typically a cubic polynomial in time is sufficient to capture the baseline hazard.

Regime (State) Switching Models

- When regimes (or states) are observed and persist over time, it is relatively straightforward to estimate a switching model.
- In the two-regime case, we use logit or probit to model the probability of transitioning from one regime to the the other, or more generally $\Pr(S_t = 1|S_{t-1}, X_{t-1})$.
- On the right-hand side, the regression includes the regime at time $t-1$, covariates, and covariates interacted with the regime at time $t-1$, for example, in the simplest case

$$y*_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 x_{t-1} + \beta_3 S_{t-1} \times x_{t-1},$$

where $y*_t$ is a latent variable that determines the probability of observing a one at time t.

Regime (State) Switching Models

$$y*_t = \beta_0 + \beta_1 S_{t-1} + \beta_2 x_{t-1} + \beta_3 S_{t-1} \times x_{t-1},$$

In the case of logit,

- The probability of persisting in the regime coded zero P_{00} is

$$1 - \frac{1}{1 + \exp^{-(\beta_0 + \beta_2 x_{t-1})}}.$$

- The probability of switching to the regime coded one is $1 - P_{00}$.
- The probability of persisting in the regime coded one P_{11} is

$$\frac{1}{1 + \exp^{-([\beta_0 + \beta_1] + [\beta_2 + \beta_3] x_{t-1})}}.$$

- The probability of switching to the regime coded zero is $1 - P_{11}$.

On space: Spatial Filtering

However, similar evasions are not available with spatial dependence. . .

On space: Spatial Filtering

However, similar evasions are not available with spatial dependence. . .

But! Spatial filtering is extremely flexible and can be used here as well. Much better alternative than using the spatially lagged observed value.

On heterogeneity: RE vs. FE

Many of the same considerations as with interval-valued outcomes (i.e., heterogeneity in the intercepts, orthogonality of the unit effects and the predictors, distribution of the unit effects, etc. . .).

On heterogeneity: RE vs. FE

Many of the same considerations as with interval-valued outcomes (i.e., heterogeneity in the intercepts, orthogonality of the unit effects and the predictors, distribution of the unit effects, etc. . .).

Some additional complications with fixed effects however:

- Cannot demean the data for fixed effects (therefore, incidental parameters problem)
- Unconditional vs. Conditional
- Sample Censoring

On heterogeneity: PML-FE

Consequence of sample censoring:

$$\underbrace{\Pr(Y=1)}_{\text{Population Relationship}} = \underbrace{\Pr(Y=1|Z=1)}_{\text{Uncensored Relationship}} \underbrace{\Pr(Z=1)}_{\text{Probability Uncensored}} + \underbrace{\Pr(Y=1|Z=0)}_{\text{Censored Relationship}}(1-\Pr(Z=1))$$

$$\text{AME} = \frac{1}{n}\sum_{i=1}^{n} \Pr(Y \widehat{= 1|x_i}, \hat{\beta}, \hat{\alpha}_i) \times (1 - \Pr(Y \widehat{= 1|x_i}, \hat{\beta}, \hat{\alpha}_i)) \times \hat{\beta}$$

On heterogeneity: PML-FE

- Penalized Maximum Likelihood due to Firth (1993)

$$L^*(\theta) = L(\theta)|I(\theta)|^{\frac{1}{2}}.$$

- Shown to be a solution for separation (Heinze and Schempher 2001)
- Sample censoring induced from unconditional fixed effects is separation

Therefore you can using PML to estimate a fixed effects model with binary outcomes (Cook, Hays, Franzese 2018)

In Sum

So often our recommendation for a ‘simple’ model with binary-TSCS data would be a PML-FE model with cubic polynomials and spatial filtering.