Spatial Models

Gregory Wawro

POLS W4292

Spatial dimensions

- Spatial interconnectivity is ubiquitous; standard & even more complex assumptions of independence are wrong & lead to incorrect inferences
- Lots of work in social sciences—environmental, urban/regional, and real-estate economics; network analysis in sociology.
- In Political Science, Franzese and Hayes have led the way, calling for reconsidering approach of almost all observational studies.
- Policy diffusion in states; increasing sensitivity due to globalization.

Concerns

- Key concern: complex patterns of correlation among outcomes due to spatial interconnectivity (not just about geography, tho).
- Can have spatio-temporal interdependence as well.

General issues

- How to deal w/ massive endogeneity in observational data (outcome for *i* explains outcome for *j* $\forall$ *i* & *j*)?
- Nuisance v. substance?
- If "nuisance": how to model relations among unobservables?
- If go "substance" route (i.e., explicitly model lags), how to come up w/ tractable estimator w/ desirable properties, esp. in small samples?

Specific Issues

- Common exposure &/or contagion $\rightarrow$ spatial association
  - **Common exposure**: units may be responding similarly to similar exposure to similar exogenous internal/domestic or external/foreign stimuli
    Contagion: responses may depend on others' responses
- Tobler's law: Everything is related to everything else, but near things are more related than distant things.
- Gives us some hope for identification: can place some structure on problem.

Spatial Autoregressive Models

- **Spatial lag models**: specify a matrix $\mathbf{W}$ that describes relationship among units & estimate ρ that gives general strength of that relationship, conditioned by $\mathbf{W}$.
- $\mathbf{W}$ is an $n \times n$ spatial-weights or connectivity matrix & ρ captures the strength of interdependence by pattern given in $\mathbf{W}$.
- $\mathbf{Wy}$ is a spatial lag. Interested in models like

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\beta + \varepsilon$$

  where the interdependence is captured by $\rho \sum_{j \neq i} w_{ij} y_j$.
- Gives reduced form

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon$$

  where $(\mathbf{I} - \rho\mathbf{W})^{-1}$ is the spatial multiplier.

Spatial Autoregressive Models

- **Spatial error models**: dependence attributable to unmeasured covariates.
- **Spatial lag and errors models**: combo (cf. dynamic panel models)

Some options for estimation

- Plain, old OLS.
- Spatial OLS: regress $\mathbf{y}$ on $\mathbf{Wy}$.
  - Inconsistent (need $\mathbf{Wy} \perp \varepsilon$), but may not do too badly. Induces exaggerated estimates of ρ (interdependence strength) and correspondingly deflates estimates of β the importance of nonspatial factors.
  - Ignoring simultaneity bias better than doing plain OLS.

Estimation options

- Spatial-2SLS/Spatial-GMM: generate spatial lags of $\mathbf{X}$ using $\mathbf{WX}$, which serves as instruments for $\mathbf{Wy}$.
    - Need truly exogenous $\mathbf{X}$.
    - Works well, even with small samples, but inefficient (worse than S-OLS).
- Spatial-ML: estimate likelihood function
    - Computat'ly intense & and doesn't appear to give efficiency gains that are significant enough to justify expense.

Spatial dependence in parameters

- An alternative to spatial lag models is to allow for interdependence in parameter values across units.
- GAM: $y_i = \alpha + \sum_j f_j(X_j) + \varepsilon_i$, where an additive predictor replaces the linear predictor in GLM.
- STructured Additive Regression (STAR) models: more flexible than GAMs—one or two dimensional nonlinear effects of continuous covariates, time scales plus unit- or cluster-specific and spatial heterogeneity.
- Combine with Bayesian approach where priors capture temporal/spatial dependence (smoothing).
- Directed/undirected/Markov random field (MRF) priors—smoothing over time/space—let periodization/contextualization schemes be substantially data-dependent.

STAR model example

$$\eta_{ijkm} = f_1(\alpha_{jk}) + f_2(UN_{ijk}) + f_3(URB_{ijk}) + f_4(AA_{ijk}) + \delta' \mathbf{z}_{ijk} \qquad (1)$$

- Use probit link fcn to model prob. senator i from region k in Congress/period j votes the pro-labor position on rollcall m.
- $f()$: smooth fcns of covariates that permit their effects to vary over region and time period.
- Can capture temporal component of accelerating concern (periodicity) & context unique to specific vote (e.g., WWII), while allowing for dependence across region.

Priors

MRF priors—e.g., for the region-period effect:

$$f(\alpha_{jk})|f(\alpha_{hl}),\ jk \neq hl,\ \tau^2 \sim N\left\{\sum_{jk\in\partial_{hl}} f(\alpha_{jk})/n_{jk}, \tau^2/n_{jk}\right\} \qquad (2)$$

- n_{jk} is # of adjacent sites & $jk \in \partial_{hl}$ denotes that site jk is a neighbor of site hl.
- Map that takes the form of a 3×8 table (3 regions over 8 congresses $\rightarrow$ parameters smoothing according to adjacency by period and region.

Extensions

- Extend to models with qualitative/limited dependent variables.
- Increasingly complex spatio-temoral patterns.

Software

- `BUGS`: very user-unfriendly and useful only for simplest of models.
- `BayesX` is user friendly for a number of models, but black-box-y and not flexible for extensions.
- `Stan` is less user-friendly, but very flexible. Code library in progress for estimating a variety of models.