Classification: A Brief Introduction to Limited Dependent Variable Spatial Models

Rao Sihang 2019.04.12

Introduction to spatial model

In reality, many variables and data have spatial dependence, for example, students' cognitive outcomes(scores) are likely to be influenced by his or her peers' outcomes, and conversely, the student's outcomes tend to his or her peers' outcomes. In that case, in order to set a moderl illustrating the correalation from those factors impacting student's outcomes, it might necessary for us to use spatial model to capture this spatial dependence with a spatial lag of the dependent variable.

Generally, there are three main kind of spatial models:

- Dependent variable y of unit A ↔ Dependent variable y of unit B

Dependent variable of a particular unit A depends on the dependent variable of other units, among which, say, unit B, and vice versa (like the case of students' cognitive outcomes).

- Independent variable x of unit B↔ Dependent variable y of unit A

Dependent variable of a particular unit depends on independent explanatory variables of other units, for example, the construction of public transportation in suburban district might have influence on the housing price of central city.

- Error term u of unit A ↔ Error term u of unit B

Determinants of the dependent variable omitted from the model are spatially autocorrelated.

A full model with all three types of interaction effects takes the form :

$$Y = \alpha l_n + \rho WY + \beta X + \theta WX + U$$
$$U = \gamma WU + \varepsilon$$

W is a nonnegative N×N(n is the number of observations) matrix describing the spatial configuration or arrangement of the units in the sample.

Limited Dependent Variable Spatial Models[1]

If the dependent variable is binary(vote or not vote) or multinomial(individual choice among different brands), we call these models "limited dependent variable spatial models.

[1]The paper, "Spatial discrete choice and spatial limited dependent variable models:a review with an emphasis on the use in regional health economics." introduce the development of limited dependent variable spatial models and the application in empirial analyses in regional health economics.

SAR Probit Model

Binary dependent variable

Firstly, I want to introduce how to treat latent variable through Beyesian based on SAR(spatial autoregression model) with binary dependent variable. The form of SAR model is shown below(only dependent variable of different observations have correlation):

$$Y = \alpha l_n + \rho WY + X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 l_n)$$
$$\Longrightarrow (I - \rho W)Y = \alpha l_n + X\beta + \varepsilon$$
$$\Longrightarrow Y = (I - W)^{-1}\alpha l_n + (I - W)^{-1}X\beta + (I - W)^{-1}\varepsilon$$

Because the dependent variable is binary with value 0 or 1, it is not easy to directly estimate the parameters.And Beyesian approach take this binary variable as the proxy of those unobservable latent utility. For example, in the context of decision on vote, if the net utility from voting is positive,you decide to vote(y=1), but if the net utility from voting is negative,you won't vote(y=0).We can use $U_1$and U_2 to denote the utilities of two choices(vote or not vote) respectively and use$y^* = U_1 - U_0$, this continuous variable to replace the binary dependent.

$$y = \begin{cases} 1 & y^* < 0 \\ 0 & y^* > 0 \end{cases}$$

Then, the model is transferred like this:

$$Y^* = (I - \rho W)^{-1}\alpha l_n + (I - \rho W)^{-1}X\beta + (I - \rho W)^{-1}\varepsilon$$

And the distribution of $p(y^*|\beta, \rho, y)$ is truncated multivariate normal distribution with mean$(I - \rho W)^{-1}X\beta$ and variance$[(I - \rho W)'(I - \rho W)]^{-1}$.

According to Albert and Chib (1993), $P(\beta, \rho|y^*) = P(\beta, \rho|y^*, y)$, so we can use the new continuous variable $y^*$to estimate the values of parameters.(Because we assume the continuous variable is truncated multivariate normal distribution,so the distribution of parametes are relatively less complicated, which simplify the parameter estimation.)

Multinomial and ordereddependent variable

Similarly, if the dependent variable is multinomial and ordered, then we can replace the discrete variables into continuous variable to do estimation.For example, in terms of the opinion on a specific statement,“ I don't like watching TV”, there are six scores representing different attitudes, 1 means extremely agree with it, and 6 means extremely disagree with it, in this case, the discrete variable, the attitude towards watching TV is ordered. However, in the context of spatial model, it is not a simple thing to directly estimate the papameters and based on Beyesian method, we use a continuous variable $y_j^*$to replace it.

$$y_j^* = U_j - U_0, \phi_{j-1} < y_j^* < \phi_j \to y = j$$

SAR Tobit Model

When the dependent variable is non-nagative(some observations with zero value,for example,the individually weekly expenditure on dining out), usually, we resort to Tobit model. And if the dependent variable has spatial dependence, we can use SAR Tobit Model, and the parameter estimation is the similar, but we only need to generate latent variable$y^*$for censored observarions(those person never dine out).

$$Y^* = (I - \rho W)^{-1} \alpha l_n + (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$$

$$y^* = \begin{cases} y_1^* & , y* \leq 0 \\ y_2 & , y* > 0 \end{cases}$$

Bibliography

[1] James LeSage and R. Kelley Pace. Introduction to spatial econometrics.

[2] J. Paul Elhorst.Spatial Econometrics From Cross-Sectional Data to Spatial Panels.