Essays on Multivariate and Simultaneous Equation Spatial Autoregressive Models

Panel databases with cross-sectional dependent variables have highlighted the need for new data analysis techniques to model dependence patterns within these databases. Among various models that describe cross-sectional dependence, the spatial autoregressive (SAR) model has attracted much attention. Since a large portion of economic theory may be concerned with interrelations among several economic variables, my research focuses on econometric studies regarding multivariate and simultaneous equations SAR models.

My job market paper, "**Multivariate and Simultaneous Equation Dynamic Panel SAR Models: Stability and Spatial Cointegration**" with Lung-fei Lee, investigates identification, estimation and application of multivariate dynamic models with spatial interaction. In addition to the effects of time lags, a cross-sectional unit may respond to its neighbor or peers' behavior in the current period (spatial lags), and in previous periods (diffusion). Therefore, a shock to one variable in one region could propagate to other variables, across dimensions of time and space. For instance, in the empirical application, we study dynamics of rice and wheat prices for 49 years of 65 cities in Yangtze River basin. One can expect that the wheat price in a city responds to the changes in rice and wheat prices in neighboring cities in the current period and previous periods, in addition to prices in the city in last period. In order to fit more empirical specifications, overall disturbances may include time fixed effects and individual fixed effects in addition to idiosyncratic shocks. We investigate two cases: all the time series are stationary, and some of them are non-stationary. For the latter, we might have cointegration across variables and spatial cointegration across spatial units.

The identification of the simultaneous equation dynamic panel SAR model follows the same strategy that we introduced in the static model paper, by providing two sets of conditions in terms of the existence of instruments for spatial lags and regularities of the weight matrix structure to obtain information from disturbance terms. We employ Quasi-ML to estimate the model under the sample scenario that both n and T tend to infinity. The time fixed effects are eliminated first and the individual fixed effects are concentrated out. The QMLEs are asymptotically normal or converge to a degenerate distribution, with asymptotic bias occurring, depending on the ratio of n and T. If there are unstable components in the time series, there may exist linear combinations among variables and across spatial units with stationary errors. We analyze conditions under which cointegration exists. Cointegration can be revealed in our model via both as an error correction representation and as a unit roots process formed by rotation of variables. The unit roots representation via rotation also helps the determination of the cointegration rank. For estimation, the QMLEs may have different rates of convergence rather than the usual root-nT-rate. While individual coefficient estimates are still root-nT consistent, the estimate of a relevant linear combination of coefficients can be superconsistent.

We apply the model with spatial cointegration to study the grain market integration using a unique historical panel dataset of grain prices from China. Previous studies only consider rice prices, while we add the multivariate feature since wheat is believed to be a substitute for rice. The empirical results show that the rice prices and wheat prices are spatially cointegrated among each other across cities. These results provide evidence of interregional and intertemporal grain market integration and the existence of a trading network in the eighteenth-century Yangtze River basin.

The second paper, "**Identification and QML Estimation of Multivariate and Simultaneous SAR Models**", with Lung-fei Lee, investigates a simultaneous equation SAR model which consists of a finite number of equations, and incorporates simultaneous effects, own-variable and cross-variable spatial lags as explanatory variables. The presence of dependent variables as endogenous regressors raises the problem of identification of simultaneous effects, while the presence of spatial lags raises the question of identification of spatial interactions. Previously suggested analyses tend to be difficult when there are more than two dependent variables. We treat the implied multivariate SAR model, which differs from the former model by excluding simultaneous effects, as a "quasi-reduced" form of a simultaneous equation model. The identification from quasi-reduced form parameters to structural parameters is similar to a traditional linear simultaneous equation model. However, the identification of a multivariate SAR model is not trivial, so we provide two sets of identification conditions in terms of the existence of instruments for spatial lags and regularities of the weight matrix structure. We study asymptotic properties of QMLEs. Monte Carlo experiments illustrate the QMLEs have broader applicability and efficiency improvement, compared to previously suggested IV-based estimates.