DSSR

Discussion Paper No.98

A threshold extension of spatial dynamic panel model

Junyue Wu
Yasumasa Matsuda

July, 2019

Data Science and Service Research Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

A threshold extension of spatial dynamic panel model

Junyue Wu[a], Yasumasa Matsuda[a]

[a] *Graduate School of Economics and Management, Tohoku University.*

Abstract

This paper proposes a threshold extension of Spatial Dynamic Panel Data (SDPD) model with fixed two-way effect to analyze data sets with spatial-temporal heterogeneity. We classify multiple regimes by a threshold variable to examine the regional dependency of parameters in SDPD models. A Bayesian estimation method along with a maximum likelihood one is put forward and compared by their Monte Carlo performance results. We find out that our Bayesian method yields more preferable estimation results, though at the expense of computation time. We also illustrate empirical applications of the threshold SDPD model to two spatial panel data set, US cigar demand data from 1963 to 1992 and Japan foreign labour data from 2008 to 2014, showing that meaningful regional dependencies of SDPD model parameters were detected.

Keywords: spatial, dynamic panel, threshold, Bayesian

1. Introduction

Panel data model with spatial interactions have been increasingly attracting attention since Anselin (1988) in the field of econometrics. A number of different settings, such as model with spatial lag or spatial error, static or dynamic, fixed or random individual effect, have been explored and their corresponding estimation methods were established , and Spatial Dynamic Panel Data (SDPD) model with two-way effect is one of the most popular models in spatial panel data analysis (Lee and Yu, 2010a). Previous researches in this field seem to favour random effect models(Baltagi et al., 2013; LeSage and Chih, 2018), however, it is often ignored whether the random effect is an appropriate specification or not. Generally, its fixed effect counterpart is more suitable in the field of spatial econometrics (Elhorst, 2014, Chapter 3).

Recently, regional heterogeneity has been introduced for spatial panel data models (Aquaro et al., 2015). Threshold models often work to account for temporal or regional heterogeneity of model parameters. Tong pioneered the threshold model in the time-series literature by proposing the self-exciting threshold autoregressive (SETAR) model and later generalized to the threshold autoregressive(TAR) model (Tong and Lim, 1980). Hansen (1999) introduced the threshold technique to the field of panel data, providing estimation and testing method for non-dynamic threshold panels . Unfortunately, threshold models only have limited presence in spatial panel literature. Majumdar et al. (2005) et.al proposed a spatio-temporal model which allows certain parameters to shift at a given time point, namely temporal dependency of parameters.

We propose a threshold extension of SDPD model with fixed two-way effect to detect the regional dependencies of parameters, specifically when the model is split into multiple regimes by a threshold

Email addresses: wu.junyue.p1@dc.tohoku.ac.jp (Junyue Wu), yasumasa.matsuda.a4@tohoku.ac.jp (Yasumasa Matsuda)

variable. Both maximum likelihood and Bayesian methods are provided for the model estimation. A comparison by Bayesian versus maximum likelihood estimation, which is uncommon in previous research, is conducted in this study. Under fair conditions, our Monte Carlo experiment reveals differences in performance and speed between the two estimation methods.

Two empirical illustrations based on panel data sets for US cigarette demand and foreign labour in Japan are used to demonstrate the situation where the proposed model succeeds in identifying the regional dependencies of parameters on a threshold variable. The former data set has been extensively examined in the panel data literature(Baltagi and Levin, 1986; Baltagi and Li, 2004) and our result is consistent with the results in the previous papers with interesting new findings.

In section 2 we set forth a threshold SDPD model along with time dependency. The maximum likelihood estimation method is described in section 3, and its Bayesian counterpart in section 4. Section 5 provides the setup and results of our Monte Carlo simulations to compare the performance of the estimation methods. Section 6 presents two empirical analysis using our proposed model, while section 7 offers our conclusion.

2. Model

Let y_{ti} be a spatial panel data at time t and regional unit i for $t = 1, \ldots, T$ and $i = 1, \ldots, N$. We consider a spatial dynamic model with fixed two-way effect as our base model, which can be regarded as a reduced version of the general spatial panel model summarized by Elhorst (2014). Let W be a predetermined $N \times N$ spatial weight matrix whose (i, j)th element represents an economic distance between the units i and j, where diagonal elements are zero and row sums are normalized to be one. Then our base model is described by

$$\begin{aligned} y_{ti} &= a_i + b_t + c + x_{ti}\beta + \theta y_{t-1,i} + \rho \sum_{j=1}^{N} w_{ij} y_{ti} + \lambda \sum_{j=1}^{N} w_{ij} y_{t-1,i} + u_{ti}, \\ u_{ti} &= \alpha \sum_{j=1}^{N} w_{ij} u_{ti} + \varepsilon_{ti}, \\ \varepsilon_{ti} &\sim N(0, \sigma^2), \end{aligned} \tag{1}$$

where a_i and b_t are two way fixed effects of individual and trend terms, c is an intercept, x_{ti} is an $1 \times H$ vector of explanatory variables and ρ, α, λ and θ are all scalar parameters reflecting the strength of spatio-temporal dependence.The spatial dynamic panel model with two way fixed effects has been studied extensively in the literature both theoretically and empirically (Lee and Yu, 2010a).

Some empirical studies pointed out that regional heterogeneity of the regression parameters β as well as the spatio-temporal ones of θ, ρ, λ and α were often detected and suitable modeling for the heterogeneity could significantly improve the model performance (LeSage and Chih, 2018). Here we propose a threshold extension of the aforementioned base model as a candidate to account for the regional heterogeneity. It is inspired by threshold models in time series literature (Tong and Lim, 1980) where model switches by lagged dependent variable. In spite of popularity in time series literature, threshold models have rarely been utilized in the field of spatial econometrics. Hansen (1999) proposed a threshold non-dynamic panel with fixed individual effects, allowing parameters for regressors x_{ti} to switch between two regimes. Majumdar et al. (2005) proposed a spatio-temporal model with a mean shift at certain time points.

Let z_{ti} be a threshold variable, which can be lagged dependent variable $y_{t-1,i}$, regressors x_{ti}, or other exogenous variables. $R_1, R_2, ..., R_q, ..., R_Q$ are subsets of real numbers, and

$$R_1 \cup R_2 \cup ... \cup R_Q = \mathbb{R}.$$

Depending on which R_q that z_{ti} falls into, we split the model into Q regimes. Then the model can be described as

$$
\begin{aligned}
&y_{ti} = a_i + b_t + \sum_{q=1}^{Q}(c_q + x_{ti}\beta_q + \theta_q y_{t-1,i} + \rho_q \sum_{j=1}^{N} w_{ij}y_{tj} + \lambda_q \sum_{j=1}^{N} w_{ij}y_{t-1,i})I\{z_{ti} \in R_q\} + u_{ti}, \\
&u_{ti} = \sum_{q=1}^{Q}(\alpha_q \sum_{j=1}^{N} w_{ij}u_{tj})I\{z_{ti} \in R_q\} + \varepsilon_{ti}, \\
&\varepsilon_{ti} \sim N(0, \sigma^2),
\end{aligned} \tag{2}
$$

where c_q, ρ_q, α_q, θ_q and λ_q are the parameters in the qth regime. In comparison with the existing approaches by Hansen (1999) and Majumdar et al. (2005), our model is more flexible in the sense that it allows for parameters in more general dynamic panel models to be dependent jointly on time t and regional units i.

To write the model in matrix notation, first we denote

$$
\begin{aligned}
x_{q,ti} &= \begin{cases} [1, x_{ti}], & \text{when } \{z_{ti} \in R_q\}, \\ 0_H, & \text{when } \{z_{ti} \notin R_q\}, \end{cases} & X^* &= [X_1, X_2 ... X_q, V], \\
\beta^* &= \begin{bmatrix} c_1 \\ \beta_1 \\ \cdots \\ c_Q \\ \beta_Q \\ a_i \\ b_t \end{bmatrix}, & D &= I_T \otimes W,
\end{aligned} \tag{3}
$$

where X_j is an $NT \times H$ matrix where $x_{q,ti}$ stacked over t and i, V is an $NT \times (N + T - 2)$ dummy matrix corresponding to $\left[\begin{smallmatrix} a_i \\ b_t \end{smallmatrix}\right]$. Then we combine each two matching parameters into one matrix

$$
\begin{aligned}
b_{ti}^{\rho} &= \rho_q d_{ti}, \quad \text{when } \{z_{ti} \in R_q\}, & b_{ti}^{\alpha} &= \alpha_q d_{ti}, \quad \text{when } \{z_{ti} \in R_q\}, \\
b_{ti}^{\lambda} &= \lambda_q d_{ti}, \quad \text{when } \{z_{ti} \in R_q\}, & d_{ti,ti}^{\theta} &= \theta_q, \quad \text{when } \{z_{ti} \in R_q\},
\end{aligned} \tag{4}
$$

where d_{ti} is the tith row of D. Then using (3) and (4) the model can be expressed as

$$
\begin{aligned}
Y &= B^{\rho}Y + (B^{\lambda} + D^{\theta})Y_{-1} + X^*\beta^* + u, \\
u &= B^{\alpha}u + \varepsilon,
\end{aligned} \tag{5}
$$

where Y is a $NT \times 1$ vector of dependent variable, Y_{-1} the lagged dependent variable, B^{ρ}, B^{λ}, B^{α} and D^{θ} are matrices constructed by stacking b_{ti}^{ρ}, b_{ti}^{α}, b_{ti}^{λ} and $d_{ti,ti}^{\theta}$ respectively.

3. Estimation

Estimation methods have been developed for spatial dynamic panel models. Yu et al. (2008) proposed a bias-corrected Quasi-Maximum Likelihood (QML) estimation for a spatial autoregressive dynamic panel model with fixed individual effect, and later Lee and Yu (2010b) expanded it to a two-way fixed effect model. Parent and LeSage (2011) constructed a space–time filter to estimate spatial panel with random effect using Bayesian Markov chain Monte Carlo (MCMC) methods.

In this section, we introduce Bayesian estimation as well as classical ML estimation to identify the proposed threshold model. We consider both estimation methods since existing studies revealed ML is prone to bias when time length is short and the number of regional units are large (Elhorst, 2014, Chapter 4) while Bayesian ones might be less susceptible.

Threshold variable z_{ti} and all the corresponding R_j are assumed to be fixed and known prior to estimation in this paper, because (i) our focus is on showing difference (in βs, for example) among regimes rather than finding the optimal threshold in the sense of goodness of fit, (ii) estimating threshold value like other parameters may bring computation issues (Hansen, 1999) . In practical situations when they are unknown, we select the combination of z_{ti} and R_js that reveals meaningful differences in regimes from predetermined candidates in a trial and error fashion.

3.1. Maximum likelihood estimation

The initial value for the dependent variable(y_{0i}) is treated as exogenous in this study. Previous study showed that initial observations without supposing any model may cause poor performances especially when T is small(Parent and LeSage, 2011), and endogenous specifications have been proposed in spatial panel literature using the approximations by Bhargava and Sargan (1983) and Elhorst (2001). But since their method can not be applied to our threshold model in a straightforward way, we stick to the exogenous specification.

Given

$$\Omega_\rho = I_{NT} - B^\rho, \qquad \Omega_\alpha = I_{NT} - B^\alpha, \tag{6}$$

and using the fore-mentioned matrix notation (5), the log-likelihood function takes the form:

$$\begin{aligned} \log \ell = & -\frac{NT}{2}\log(2\pi\sigma^2) + \log|\Omega_\rho| + \log|\Omega_\alpha| \\ & - \frac{1}{2\sigma^2} K^{*T} {\Omega_\alpha}^T \Omega_\alpha K^*, \\ K^* = & \, \Omega_\rho Y - (B^\lambda + D^\theta) Y_{-1} - X^*\beta^*, \end{aligned} \tag{7}$$

To improve the efficiency of the calculation, β^* and σ^2 can be concentrated out as

$$\begin{aligned} \hat{\beta^*} &= (X^{*T}{\Omega_\alpha}^T\Omega_\alpha X^*)^{-1} X^{*T}{\Omega_\alpha}^T\Omega_\alpha(\Omega_\rho Y - (B^\lambda + D^\theta)Y_{-1}), \\ \hat{\sigma^2} &= \frac{1}{NT}(\Omega_\rho Y - (B^\lambda + D^\theta)Y_{-1} - X^*\hat{\beta^*})^T {\Omega_\alpha}^T \Omega_\alpha (\Omega_\rho Y - (B^\lambda + D^\theta)Y_{-1} - X^*\hat{\beta^*}). \end{aligned} \tag{8}$$

Substituting (8) into (7) leads to the concentrated log-likelihood function, in which only ρ_j, α_j and λ_j need to be estimated

$$\log \ell_c = -\frac{NT}{2}\log(2\pi\hat{\sigma^2}) + \log|\Omega_\rho| + \log|\Omega_\alpha|.$$

In the case that the threshold variable is not time dependent and remains constant over all the time period, $\log|\Omega_\rho| + \log|\Omega_\alpha|$ can be replaced by $T\log|I_N - \widetilde{B}^\rho| + T\log|I_N - \widetilde{B}^\alpha|$ to speed up the calculations, where

$$\widetilde{B}^\rho_{i,j} = \rho_j W_{i,j}, \quad \text{when } \{z_{ti} \in R_j\}, \qquad \widetilde{B}^\alpha_{i,j} = \alpha_j W_{i,j}, \quad \text{when } \{z_{ti} \in R_j\}.$$

The reduction of computational time by the substitution in the case will hold as well for Bayesian method in the next section, since the Bayesian method employs the log-likelihood evaluation in their estimation procedure.

It should be notified that the demeaning technique usually employed to eliminate the cross-sectional fixed effects a_i does not work for the threshold model, since $QB^\rho Y \neq B^\rho QY$ and $QB^\alpha Y \neq B^\alpha QY$ unless its threshold variable is time independent, where

$$\begin{aligned} Q =& (I_T - \frac{1}{T}L_T) \otimes I_N, \\ \overline{Y} =& QY, \end{aligned}$$

where L_T is a $T \times T$ matrix of ones.

Finally, stationary conditions for the model are satisfied when both $|\Omega_\alpha^{-1}|$ and $|(D^\theta + B^\lambda)\Omega_\rho{}^{-1}|$ are less than 1 (Elhorst, 2014). This condition should be satisfied during the optimization process for ML.

3.2. Bayesian estimation

We take the approach illustrated by Tsay (2005) to implement our MCMC method. First, we need to specify prior distributions for all the parameters. We use conjugate prior distributions to obtain closed-form expressions for the conditional posterior distributions necessarily when they are available. Using the notations from the previous section, we specify normal prior for β^* and inverse Gamma prior for σ^2 by

$$\beta^* \sim N(M_{\beta^*}, T_{\beta^*}^{-1}), \qquad \sigma^2 \sim IG(G_a, G_b), \tag{9}$$

where M_{β^*} is set to be a $(Q(H+1)+2) \times 1$ vector of zeros, T_β^* a $(Q(H+1)+2) \times (Q(H+1)+2)$ matrix of zeros, and $G_a = G_b = 0$ in the following sections to conduct a fair comparison with the ML method. Multiplying the priors to the likelihood function in the previous section, we obtain the conditional posterior distributions for β^* and σ^2 by the Bayes theorem in the closed form as

$$\begin{aligned}
(\beta^*|Y, X^*, \Omega_\rho, \Omega_\alpha, B^\lambda, D^\theta, \sigma^2) &\sim N(M_{p\beta^*}, T_{p\beta^*}^{-1}),\\
M_{p\beta^*} &= T_{p\beta^*}^{-1}(T_{\beta^*} M_{\beta^*} + \frac{1}{\sigma^2} X^{*T} \Omega_\alpha{}^T \Omega_\alpha (\Omega_\rho Y - (B^\lambda + D^\theta) Y_{-1})),\\
T_{p\beta^*} &= T_{\beta^*} + \frac{1}{\sigma^2} X^{*T} \Omega_\alpha{}^T \Omega_\alpha^* X^*,\\
(\sigma^2|\Omega_\alpha, K^*) &\sim IG(G_a + \frac{NT}{2}, G_b + \frac{1}{2} K^{*T} \Omega_\alpha{}^T \Omega_\alpha K^*).
\end{aligned} \tag{10}$$

For the priors of ρ_q, α_q, θ_q, and λ_q, we specify the uniform distribution over the regions in which the fore-mentioned stationary condition is satisfied. The conditional posteriors take the form in

$$\begin{aligned}
(\rho_q|\Omega_\alpha, \Omega_\rho, K^*, \sigma^2) &\propto exp(\frac{K^{*T}\Omega_\alpha{}^T\Omega_\alpha K^*}{2\sigma^2})||\Omega_\rho|| \qquad && q = 1, 2...Q,\\
(\alpha_q|\Omega_\alpha, K^*, \sigma^2) &\propto exp(\frac{K^{*T}\Omega_\alpha{}^T\Omega_\alpha K^*}{2\sigma^2})||\Omega_\alpha|| && q = 1, 2...Q,\\
(\theta_q|\Omega_\alpha, K^*, \sigma^2) &\propto exp(\frac{K^{*T}\Omega_\alpha{}^T\Omega_\alpha K^*}{2\sigma^2}) && q = 1, 2...Q,\\
(\lambda_q|\Omega_\alpha, K^*, \sigma^2) &\propto exp(\frac{K^{*T}\Omega_\alpha{}^T\Omega_\alpha K^*}{2\sigma^2}) && q = 1, 2...Q.
\end{aligned} \tag{11}$$

Since it does not provide closed-form posteriors, Metropolis algorithm driven by Gaussian random walk is applied to draw the posterior samples.

Combining the closed form posteriors for β^* and σ^2 with the non-closed form ones for ρ_q, α_q, θ_q, and λ_q, we propose the whole MCMC procedure as follows:

1. Specify hyperparameter values in (9).
2. Specify arbitrary (satisfy the stationary condition) starting values for all parameters.
3. Use the conditional posterior distribution in (10) via Gibbs sampling to draw a set of samples for β^* and σ^2.
4. Use the conditional posterior distribution in (11) via Metropolis algorithm to draw a set of samples for ρ_q, α_q, θ_q, and λ_q.

Repeat steps 3 and 4 to obtain posterior sample, and it can be used to calculate the estimation result.

4. Simulation Results

This section shows the conditions and results of the simulation conducted to compare the performance of the fore-mentioned maximum likelihood estimator and the Bayesian estimator. For simplicity, we only demonstrate $J = 2$ case of the threshold model (2).

We designed 50 cross-sectional units, 5 time period and 3 independent variables in (2) to simulate panel data. A small T scenario was chosen since it is more prone to bias (Elhorst, 2010). We set all the dependent variable for the initial time period y_{0i} to 0, and generated $y_t i$ for t from 1 to 100. Then we discarded the first 94 time periods to ensure the data set is not influenced by the initial values. The spatial correlation of the cross-sectional units was based on first contiguity of 5×10 grid, so each one had 2 to 4 neighbours. The exogenous explanatory variables were independently drawn from a uniform distribution on the interval $[-1, 1]$. For the disturbance in (2), we first generated ε_{ti} from i.i.d. normal distribution $N(0, 0.1)$, then took spatial correlation into account by calculating the product of Ω_α in (6) and the vector of ε. The fixed cross-sectional individual effect a_i and time effect b_t were both generated from i.i.d. normal distribution $N(0, 0.05)$. The values of all βs were predetermined and listed in the simulation results Table 1. All the other parameters for the spatio-temporal dependence were independently drawn from uniform distributions with their range listed in the table, and the combinations that do not satisfy the stationary restriction were discarded. For the threshold, we generate a vector of z_{it} from an i.i.d. standard normal distribution and divide $\mathbb{R}$ into R_1 and R_2 at the origin.

Each result shows biases and root-mean-square error from 1000 runs of the ML and Bayes estimators, where we constructed the Bayesian estimator by the mean value of the latter 8,000 posterior samples generated from 10,000 iterations. We conducted 95 % confidence interval for the regression coefficients $\beta_1, \beta_2, \beta_3$ for both ML and Bayes estimators. The estimation methods are implemented using R (R Core Team, 2016), Rcpp (Eddelbuettel and François, 2011), RcppArmadillo (Eddelbuettel and Sanderson, 2014) and RcppEigen (Bates and Eddelbuettel, 2013) library.

Table 1 reports the result of the simulation, the numbers in brackets are corresponding root-mean-square error, and the percentages that the true value falls in 95% confidence interval are listed on the right. The ML method performs similarly to the Bayesian one except for the bias in σ^2, which results in on average around 5.4% less interval estimation results contain the true value. This may cause problems when conducting t-tests for the estimation result of ML. When it comes to computation time, the ML method cost 33.1 minutes , and the Bayesian method used 153.1 minutes. Note that it is possible to cut down the computational time for the Bayesian method by doing fewer iterations, 10,000 iterations could be a bit excessive but we just want to be sure it converges.

Based on these results, we find that the Bayesian estimation reduced the negative bias for σ^2 caused by ML, while the performances over the other parameters are overall similar. It is known that Yu et al. (2008) proposed a bias reduction method for ML, but the method does not work here because of the non-linear nature and the inclusion of spatial error structure in our threshold model. The disadvantage of the Bayesian estimation is that it is more time consuming to complete the MCMC procedures to obtain the estimate, especially when the data set gets larger. One possible way to reduce the computation time is to use the restriction put forward by Yu et al. (2008), which requires $|\rho| + |\theta| + |\lambda| < 1$. This condition is overly restrictive (Elhorst, 2014, Chapter 4) but using it can avoid calculating the value of $|\Omega_\alpha|$ and $|(D^\theta + B^\lambda)\Omega_\rho{}^{-1}|$. In the following empirical studies, we stick to the original restriction because the computation time is manageable.

5. Empirical Analysis

Two empirical study were conducted to illustrate the proposed threshold model in this section. Only results for the Bayesian method is presented because of its superiority explained in last section.

Table 1: Simulation result

Parameter (= true value)	ML	in confidence interval	Bayesian	in confidence interval
$\rho_1(=0.1 \sim 0.7)$	-0.0102		-0.0287	
	(0.0581)		(0.0517)	
$\alpha_1(=0.1 \sim 0.7)$	-0.0364		0.0012	
	(0.2492)		(0.1621)	
$\theta_1(=0.1 \sim 0.7)$	-0.0249		-0.0263	
	(0.0335)		(0.0293)	
$\lambda_1(=0.1 \sim 0.3)$	-0.0177		-0.0218	
	(0.0567)		(0.0485)	
$\beta 1_1(=1.5)$	-0.0124	87.1%	-0.0212	94.3%
	(0.0574)		(0.0574)	
$\beta 2_1(=1.5)$	-0.0133	87.0%	-0.0174	93.2%
	(0.0585)		(0.0572)	
$\beta 3_1(=1.5)$	-0.0132	87.5%	-0.0171	93.6%
	(0.0590)		(0.0582)	
$\rho_2(=0.1 \sim 0.7)$	-0.0122		-0.0322	
	(0.0566)		(0.0518)	
$\alpha_2(=0.1 \sim 0.7)$	-0.0287		0.0030	
	(0.2510)		(0.1575)	
$\theta_2(=0.1 \sim 0.7)$	-0.0247		-0.0267	
	(0.0366)		(0.0293)	
$\lambda_2(=0.1 \sim 0.3)$	-0.0154		-0.0203	
	(0.0562)		(0.0496)	
$\beta 1_2(=0.5)$	-0.0036	89.4%	-0.0027	93.1%
	(0.0569)		(0.0577)	
$\beta 2_2(=0.5)$	-0.0042	88.0%	-0.0050	94.1%
	(0.0575)		(0.0562)	
$\beta 3_2(=0.5)$	-0.0029	90.3%	-0.0049	93.6%
	(0.0555)		(0.0576)	
$\sigma^2(=0.1)$	-0.0273		-0.0015	
	(0.0097)		(0.0103)	

5.1. US cigarette demand

As an applied illustration, we examined cigarette demand based on panel data from 46 American states during the time period of 1963-1992. It is a log-transformed annual data set which consists of $N = 46, T = 30$. The dependent variable is cigarette sales in packs per capita, and there are two independent variables which are per capita disposable income(DI) and price per pack of cigarettes(P). The spatial weight matrix is defined by first-order contiguity. An older version of this data set (contains data from 1963 - 1980) was first analyzed by Baltagi and Levin (1986) using a dynamic panel model, and later the updated one was revisited by Baltagi and Li (2004) with spatial correlation taken into account.

Table 3 shows the estimation result of the threshold model in (2), and its non-threshold counterpart (1) is also listed in Table 2 for comparison. The predetermined threshold variable z_{ti} is lagged disposable income $\ln INC_{t-1,i}$, and the third quartile of all $\ln INC_{ti} \quad (t = 0, 1..., T-1, i = 1, .., N)$ is used to separate the two regimes. Figure 1 shows how the threshold evolves during the 1980-1989 period, before 1980 all states are identified as low income and after 1989 all high income. For this particular data set, we found that the full model might be redundant since ρ and α were hard to achieve convergence in both threshold and non-threshold case. So we dropped ρ and fitted the reduced model. The estimations were constructed using samples drawn from 1 million iterations with the first 20% excluded for burn-in.

The non-threshold result is in line with previous studies (Baltagi and Li, 2004): (i) the presence of spatial correlation is observed ,(ii) first-order time dependency is strong, (iii) a negative relationship between prices and sales are present, (iv) positive correlation exists between disposable income and sales. Our threshold model enables us to have a closer look, A series of difference between high and low income regions reveals: (i) significant spatial correlation only exists in high-income regions, (ii) price have a larger impact in low-income regions, (iii) the impact of income is weaker in high-income regions. This result indicates that the demand of cigarette is becoming less susceptible to changes in its price and income level while disposable income is getting higher over time, which is reasonable and conforms to economic theories.

Table 2: Estimation result: non-threshold model

	coef	sd
α	0.0764	0.0380
θ	0.8249	0.0143
λ	0.0126	0.0202
c	1.2208	0.2736
$\ln P$	-0.2932	0.0242
$\ln DI$	0.1050	0.0250
σ^2	0.0012	0.0000

5.2. Foreign labour in Japan

The second example illustrates the time-independent case of our model. we examined foreign labour based on annual panel data from 47 prefectures of Japan during 2008-2014, published by Japanese Government (National Statistics Center, 2018). 3 explanatory variables were chosen: prefectural income per capita (INC), consumer price index (CPI), and the number of total foreign hotel guests(NFG) per capita. The basis of CPI is set to the year 2008 ($CPI_{2008,i} = 100$) and $CPI - 100$ is used in the analysis, other data were log-transformed.

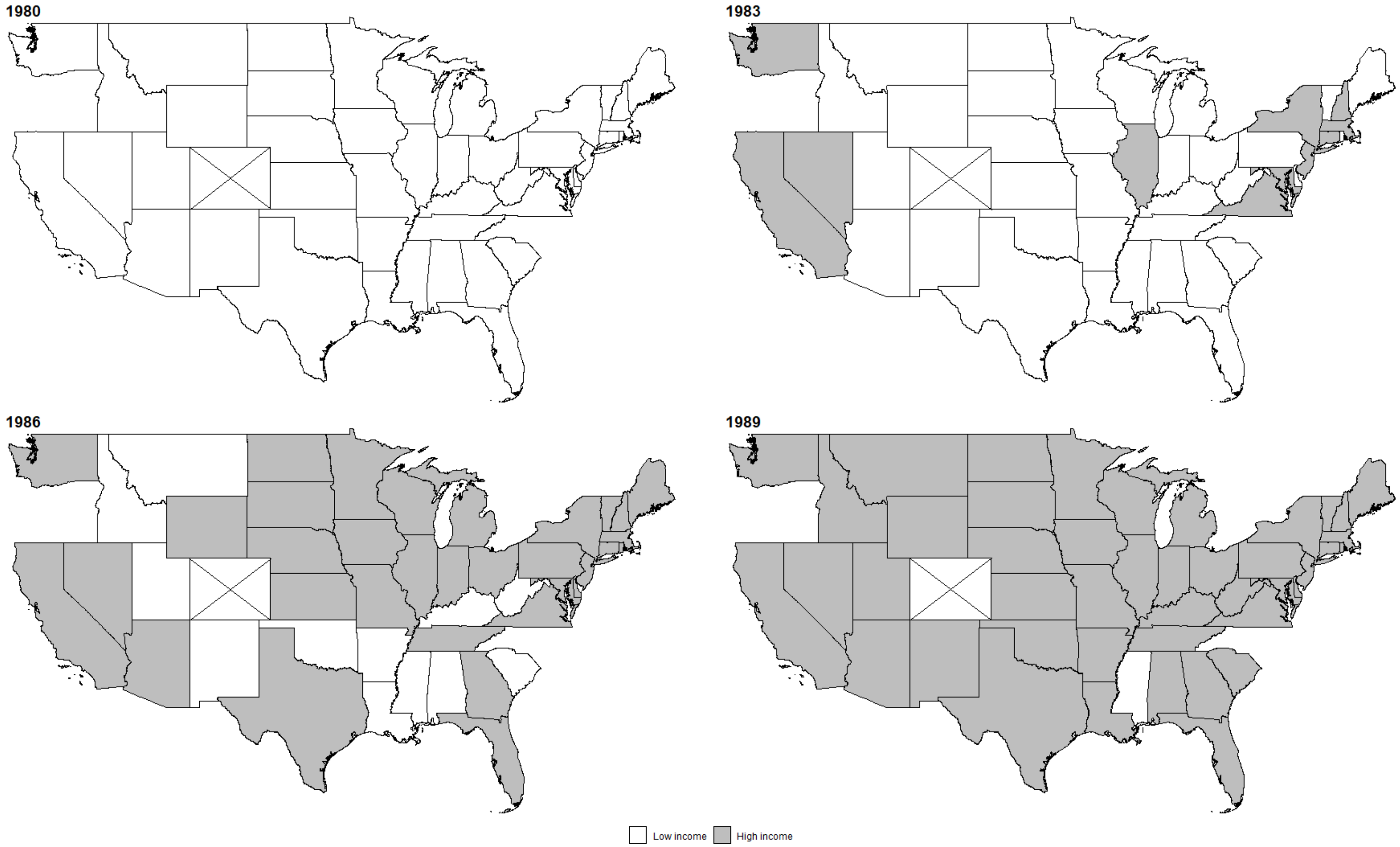


Figure 1: US States, by high and low disposable income

Table 3: Estimation result: threshold model

	coef	sd
α_1	0.1563	0.0731
θ_1	0.7731	0.0189
λ_1	0.0567	0.0255
c_1	1.7085	0.2985
$\ln P_1$	-0.2463	0.0311
$\ln DI_1$	0.0335	0.0283
α_2	0.0647	0.0485
θ_2	0.7937	0.0153
λ_2	0.0050	0.0182
c_2	0.9778	0.2853
$\ln P_2$	-0.3123	0.0242
$\ln DI_2$	0.1590	0.0270
σ^2	0.0012	0.0000

Table 5 and 4 report the estimation results of the threshold (2) and non-threshold model (1) respectively. The predetermined threshold variable z_{ti} is GPP in 2008, and the third quartile was chosen to split the two regimes (shown in Figure 2). Different from the last example, this threshold stays the same for all time periods. We encountered convergence problem similar to the last section and decided to fit the reduced model without α. The estimations were constructed using samples drawn from 5 million iterations with the first 1 million excluded for burn-in.

The non-threshold result indicates the existence of the overall spatial effect and time dependency, as well as positive impacts for income and foreign guests. The result of threshold model gives us extra insights: (i) the spatial effect is only significant for regions with lower GPP, (ii) income per capita has a significant impact in the regions where it is low. According to the annual foreign worker report published by Ministry of Health, labour and welfare (2019), most prefectures with lower GPP have around half of the foreign workers working in the manufacturing industry, on the contrary, ones with high GPP tend to have a more balanced composition of foreigners working in different industries. Also, the manufacturing industry in Japan tends to have a higher concentration of unskilled foreign workers. The result indicates unskilled foreign workers in Japan could be more sensitive to the change of income level since income has a significant impact on the prefectures where they dominate the foreign labour pool.

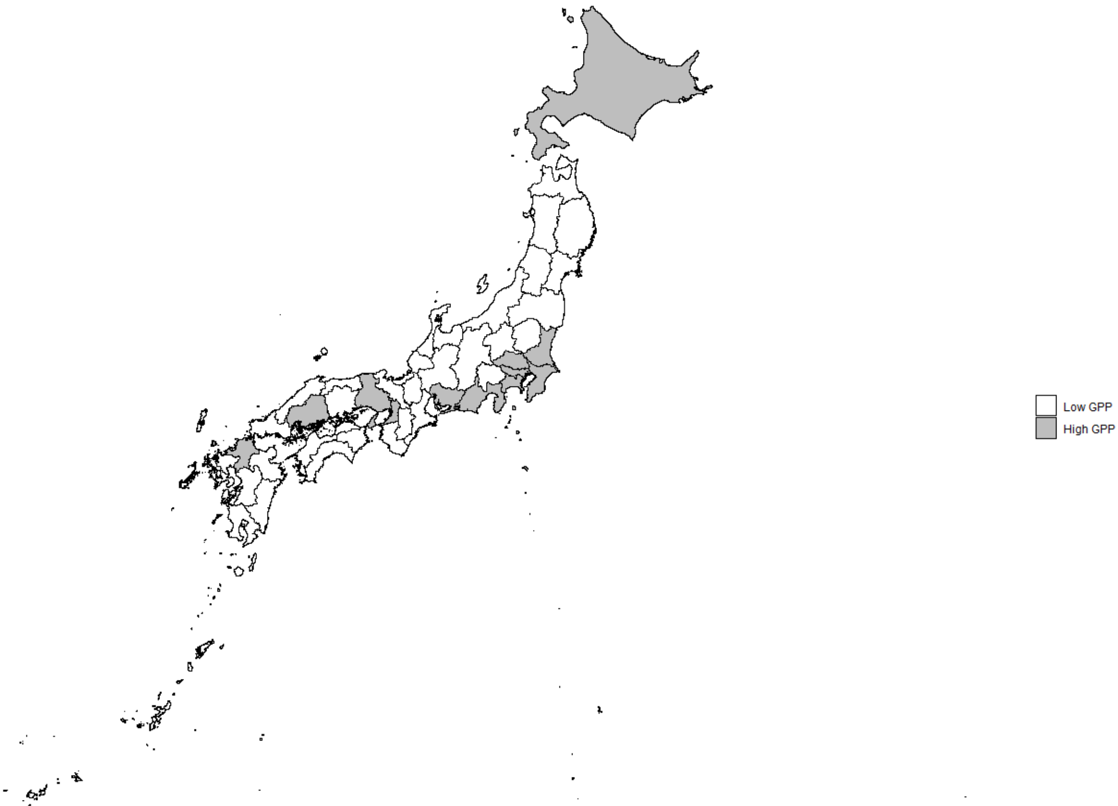
Table 4: Estimation result: non-threshold model

	coef	sd
ρ	0.1614	0.0647
θ	0.4904	0.0617
λ	-0.0258	0.0830
c	-2.6485	0.5619
$CPI - 100$	-0.0121	0.0129
$\ln INC$	0.7150	0.2082
$\ln NFG$	0.0689	0.0225
σ^2	0.0048	0.0005

Table 5: Estimation result: threshold model

	coef	sd
ρ_1	0.0940	0.1500
θ_1	0.6927	0.1150
λ_1	0.0291	0.1407
c_1	-3.4864	0.7868
$CPI - 100_1$	-0.0070	0.0144
$\ln INC_1$	-0.0196	0.4503
$\ln NFG_1$	0.0721	0.0492
ρ_2	0.2000	0.0895
θ_2	0.4332	0.0529
λ_2	-0.0413	0.0980
c_2	3.4073	7.7322
$CPI - 100_2$	-0.0206	0.0138
$\ln INC_2$	0.9567	0.2211
$\ln NFG_2$	0.0803	0.0225
σ^2	0.0047	0.0005

Figure 2: Japan Prefectures, by high and low GPP in 2008



6. Conclusion

We introduced a threshold extension of SDPD model with fixed two-way effects. This specification is most suitable when analyzing data set in which regions can be divided into groups in meaningful ways, and it can reveal differences for different kind of regions. Only model with two regimes was presented in simulation and empirical study but it could be applied to multiple regime case without many difficulties.

Bayesian and maximum likelihood estimations were set forth, and simulation experiments were conducted to compare their efficiency. The result indicates that the Bayesian method is more accurate and consistent, but it comes with the cost of computational time. We used non-informative priors for the Bayesian method to ensure the fairness of the comparison, but can be changed according to the needs of the analysis.

One remaining problem is that the threshold is determined before estimation, and it is picked by how well it presents the difference among regimes in a trial and error fashion. Since this method is time-consuming and prone to human bias, we would like to find a better solution in the future.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. volume 4. Springer Science & Business Media.

Aquaro, M., Bailey, N., Pesaran, M.H., 2015. Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. USC-INET Research Paper .

Baltagi, B.H., Egger, P., Pfaffermayr, M., 2013. A generalized spatial panel data model with random effects. Econometric Reviews 32, 650–685.

Baltagi, B.H., Levin, D., 1986. Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. The Review of Economics and Statistics , 148–155.

Baltagi, B.H., Li, D., 2004. Prediction in the panel data model with spatial correlation, in: Advances in spatial econometrics. Springer, pp. 283–295.

Bates, D., Eddelbuettel, D., 2013. Fast and elegant numerical linear algebra using the RcppEigen package. Journal of Statistical Software 52, 1–24. URL: `http://www.jstatsoft.org/v52/i05/`.

Bhargava, A., Sargan, J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods. Econometrica: Journal of the Econometric Society , 1635–1659.

Eddelbuettel, D., François, R., 2011. Rcpp: Seamless R and C++ integration. Journal of Statistical Software 40, 1–18. URL: `http://www.jstatsoft.org/v40/i08/`, doi:10.18637/jss.v040.i08.

Eddelbuettel, D., Sanderson, C., 2014. Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. Computational Statistics and Data Analysis 71, 1054–1063. URL: `http://dx.doi.org/10.1016/j.csda.2013.02.005`.

Elhorst, J.P., 2001. Dynamic models in space and time. Geographical Analysis 33, 119–140.

Elhorst, J.P., 2010. Dynamic panels with endogenous interaction effects when t is small. Regional Science and Urban Economics 40, 272–282.

Elhorst, J.P., 2014. Spatial econometrics: from cross-sectional data to spatial panels. volume 479. Springer.

Hansen, B.E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. Journal of econometrics 93, 345–368.

Lee, L.f., Yu, J., 2010a. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40, 255–271.

Lee, L.f., Yu, J., 2010b. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26, 564–597.

LeSage, J.P., Chih, Y.Y., 2018. A bayesian spatial panel model with heterogeneous coefficients. Regional Science and Urban Economics 72, 58–73.

Majumdar, A., Gelfand, A.E., Banerjee, S., 2005. Spatio-temporal change-point modeling. Journal of Statistical Planning and Inference 130, 149–166.

Ministry of Health, labour and welfare, 2019. Summary of foreigner employment situation (at the end of October,2018). URL: https://www.mhlw.go.jp/stf/newpage_03337.html.

National Statistics Center, 2018. Portal site of official statistics of japan website. URL: https://www.e-stat.go.jp.

Parent, O., LeSage, J.P., 2011. A space–time filter for panel data models containing random effects. Computational Statistics & Data Analysis 55, 475–490.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data-with discussion. Journal of the Royal Statistical Society. Series B: Statistical Methodology 42, 245–292.

Tsay, R.S., 2005. Markov chain monte carlo methods with applications, in: Analysis of financial time series. John wiley & sons. volume 543, pp. 624–628.

Yu, J., De Jong, R., Lee, L.f., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. Journal of Econometrics 146, 118–134.