Cross-sectional space-time modeling using ARNN(p, n) processes*

W. Polasek[†] K. Kakamu[‡]

September, 2006

Abstract

We suggest a new class of cross-sectional space-time models based on local AR models and nearest neighbors using distances between observations. For the estimation we use a tightness prior for prediction of regional GDP forecasts. We extend the model to the model with exogeneous variable model and hierarchical prior models. The approaches are demonstrated for a dynamic panel model for regional data in Central Europe. Finally, we find that an ARNN$(1, 3)$ model with travel time data is best selected by marginal likelihood and there the spatial correlation is usually stronger than the time correlation.

Key Words: Dynamic panel data; Hierarchical models; Marginal likelihoods; Nearest neighbors; Tightness prior; Spatial econometrics.
JEL Classification: C11, C15, C21, R11.

*We would like to thank Hedibert Freitas Lopes for his helpful comments and discussions.
[†]Department of Economics & Finance, Institute for Advanced Studies, Stumpergasse 56, 1060, Vienna, Austria, *Email*:polasek@ihs.ac.at
[‡]Department of Economics & Finance, Institute for Advanced Studies, Stumpergasse 56, 1060, Vienna, Austria, *Email*:kakamu@ihs.ac.at

1 Introduction

We propose a space-time model for predicting regional business cycle from a Bayesian point of view. Since the seminal work by Anselin (1988), spatial interaction becomes one of the concerns in economics. Therefore, the spatial dependency is modeled for several econometric models. However, the concerns are moved to space-time model (see e.g. Banerjee *et al.*, 2003).

Analyzing regional business cycles by regional models have become an important issue in recent time, since the phenomenon of non-convergence has gained more attention in the debate of regional convergence in an enlarged European Union. Therefore we approach this problem from a new econometric perspective using a new class of space-time models, the AR nearest neighbor models. Kakamu and Wago (2005) have pointed out that the spatial interaction plays an important role in regional business cycle analysis in Japan. They considered the panel probit model with spatial dependency from a Bayesian point of view and analysed the spatial interactions in regional business cycles in Japan.

The goal of this paper is to construct a model for predicting regional business cycle and to model the regional GDP dynamics of 227 regions in six countries of central Europe during the period 1995 to 2001. Furthermore, we use the concept of nearest neighbors (NN) and propose the tightness prior. Our results show that the spatial correlations are high and the serial correlations are small.

The rest of this paper is organized as follows. In Section 2, we will explain the autoregressive nearest neighbor model for regional modeling. In Section 3, wedescribe the computations by the MCMC method and the model selection procedure and extend to the one with exogenous variables and the hierarchical prior models. In Section 4, we will analyze the GDP growth in 227 regions across six countries in central Europe. Finally, some conclusions are given in Section 5.

2 Regional ARNN modeling

We consider a dynamic panel data matrix Y of order $(N \times T)$, where usually the time dimension T is much smaller than N. Let y_t denote the t-th column of Y, then we define the k-nearest neighbor matrix as $W_1 = \mathrm{NN}(1)$ until $W_n = \mathrm{NN}(n)$ where W_1 denotes the $(N \times N)$ 0-1 matrix with a 1 in each row indicating the nearest neighbor (NN) for each region, i.e. for each row. Thus, W_k denotes the matrix of the k-th nearest neighbors for each region.

2.1 Some properties of ARNN processes

Definition 1: The ARNN(p, n) processes

We consider a dynamic $N \times T$ panel data matrix and using the time lag operator L, defined by $Ly_t = y_{t-1}$ and the NN weight matrices $W_1, \cdots, W_n$ of a vectorized time series $y = \mathrm{vec}Y$ the ARNN(p, n) process is given by

$$\beta(L \circ W)y_t = u_t, \text{ for } t = 1, \cdots, T$$

where u_t, is a white noise process and the ARNN polynomial is given by

$$\beta(L \circ W) = (1 - \beta(L) \circ W) = (1 - \beta_1(L)W_1 - \cdots \beta_n(L)W_n)$$

This implies the following decomposition of the ARNN process

$$\begin{aligned} \beta(L \circ W))y_t = (1 - \beta(L) \circ W)y_t = \\ (1 - \beta_1(L)W_1 - \cdots \beta_n(L)W_n)y_t = \\ y_t - \beta_1(L)y_t^1 - \cdots \beta_n(L)y_t^n \end{aligned}$$

with $y_t^n = W_n y_t$. We define the extension of the spatial operator to include the pure AR operator.

$$\beta^0(L \circ W) = (1 - \beta^0(L) \circ W) = (1 - \beta_0(L) - \beta_1(L)W_1 - \cdots - \beta_n(L)W_n)$$

Definition 2: Stationary ARNN model

a) Stationarity condition: The ARNN(p, n) process is stationary if the pure AR(p) polynomial of the ARNN polynomial has all roots outside the unit circle.

$$\beta_0(L) = 1 - \beta_{10}L - \beta_{20}L^2 - \cdots - \beta_{p0}L^p,$$

b) The ARNN(p,n) process is called NN-stationary if the n spatial sub-processes $y_t^i = W_i y_t, i = 1, ..., n$ are also stationary and the roots of the p polynomials lie outside the unit circle:

$$\beta_k(L) = 1 - \beta_{i1}L - \beta_{i2}L^2 - \cdots - \beta_{in}L^n, \quad \text{for} \quad i = 1, \cdots, p.$$

Note that the evaluation of the ARNN polynomial follows a matrix scheme:

$$\begin{aligned}
\beta(L \circ W)y_t = (1 - \beta(L) \circ W)y_t &= (1 - \beta_1(L)W_1 - \cdots \beta_n(L)W_n)y_t \\
&= (1 - \beta_{11}LW_1 - \cdots - \beta_{1n}LW_n - \cdots \\
&\quad -\beta_{p1}L^pW_1 - \cdots - \beta_{pn}L^pW_n)y_t = u_t
\end{aligned}$$

2.2 Estimation of ARNN processes

The dependent variable is given by the most recent observed cross section column of matrix Y, i.e. $y = y_T$. Now we define a spatial AR model for each region

$$\begin{aligned}
y &= \beta_{10}y_{t-1} + \beta_{11}W_1y_{t-1} + \beta_{12}W_2y_{t-1} + \cdots + \beta_{1n}W_ny_{t-1} + \cdots \\
&\quad +\beta_{p0}y_{t-p} + \beta_{p1}W_1y_{t-p} + \beta_{p2}W_2y_{t-p} + \cdots + \beta_{pn}W_ny_{t-p} + u, \\
&= (y_{t-1}, W_1y_{t-1}, W_2y_{t-1}, \cdots, W_ny_{t-1})\beta_1 + \cdots \\
&\quad +(y_{t-p}, W_1y_{t-p}, W_2y_{t-p}, \cdots, W_ny_{t-p})\beta_p + u, \\
&= X_1^{p,n}\text{vecB} + u, \quad u \sim \mathcal{N}(0, \sigma^2 I_N), \qquad (1)
\end{aligned}$$

where the $(N \times (n+1)p)$ regressor matrix is given by

$$X_1^{p,n} = (y_{t-1}, y_{t-1}^1, \cdots, y_{t-1}^n, \cdots, y_{t-p}, y_{t-p}^1, \cdots, y_{t-p}^n), \qquad (2)$$

with $y_{t-j}^k = W_k y_{t-j}$ that is the k-th nearest neighbor of the time lag j.

And the coefficients in the columns of B, like $\beta_1 = (\beta_{10}, \cdots, \beta_{1n})'$ is the $(n+1)$-dimensional spatial AR regression vector. The whole regression coefficient matrix is now given by $(n+1) \times p$ matrix $\text{B} = (\beta_1 \cdots, \beta_p)$.

For the prior distribution of the regression coefficients we assume a tightness covariance matrix and we assume linear decreasing variance factors across the diagonal of the covariance matrix:

$$D_{in} = diag(1/i, 1/i, 1/i2, \cdots, 1/in), \qquad (3)$$

so that for each time lag i we think that the coefficients are similar and can make the same tightness distributional assumption for the regression coefficients: the i-th row vector β_i of the matrix B follows a distribution with center 0 and a variance that is closer to zero, the higher the lag order is:

$$\beta_i \sim \mathcal{N}(0, \tau_*^2 D_{in}), \text{ for } i = 1, \cdots, p \tag{4}$$

where each diagonal element of D_n decrease with larger distance, that is, a closer region can have more coefficient variation than a on than a region that is farther away.

We write the simple Bayesian ARNN(p, n) model in the compact matrix form given by

$$y = X_1^{p,n}\text{vecB} + u, \quad u \sim \mathcal{N}(0, \sigma^2 I_N). \tag{5}$$

Then, the likelihood function is as follows;

$$L(y|X_1^{p,n}, \text{vecB}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}^N} \exp\left(- \frac{e'e}{2\sigma^2} \right), \tag{6}$$

where the residuals are calculated $e = y - X_1^{p,n}\text{vecB}$ and the prior information follows a normal gamma model or is specified independently as

$$\text{vecB} \sim \mathcal{N}(0, \tau_*^2 P \otimes D_n), \ \sigma^2 \sim \mathcal{G}^{-1}(\nu_*/2, \lambda_*/2), \tag{7}$$

where $P = diag(1, 1/2, \cdots, 1/p)$ and $\mathcal{G}^{-1}(a, b)$ denotes inverse gamma distribution with parameters a and b.

The following conditions are needed to obtain a NN-stationary solution (see definition 2). The roots of the polynomials

$$\begin{gathered} 1 - \beta_{10}L - \beta_{20}L^2 - \cdots - \beta_{p0}L^p, \\ 1 - \beta_{11}L - \beta_{12}L^2 - \cdots - \beta_{1n}L^n, \\ \vdots \\ 1 - \beta_{p1}L - \beta_{p2}L^2 - \cdots - \beta_{pn}L^n, \end{gathered}$$

are all outside the unit circle.

Given the prior density $p(\text{vecB}, \sigma^2) = p(\text{vecB}|\sigma^2)p(\sigma^2)$ and the likelihood function given in (6), the joint posterior distribution can be expressed as

$$p(\text{vecB}, \sigma^2|y, X_1^{p,n}) = p(\text{vecB}, \sigma^2)L(y|\text{vecB}, \sigma^2, X_1^{p,n}). \tag{8}$$

Since the joint posterior distribution given by (8) can be simplified, we can now use MCMC methods. The Markov chain sampling scheme is constructed from the full conditional distributions of vecB and σ^2.

For vecB given σ^2, it can be easily obtained by Gibbs sampler (see Gelfand and Smith, 1990) that

$$\text{vecB}|\sigma^2, y, X_1^{p,n} \sim \mathcal{N}(\text{vecB}_{**}, \Sigma_{**}) \tag{9}$$

where $\text{vecB}_{**} = \Sigma_{**}(\sigma^{-2} X_1^{p,n'} y)$, $\Sigma_{**} = (\sigma^{-2} X_1^{p,n'} X_1^{p,n} + \Sigma_*^{-1})^{-1}$ and $\Sigma_* = \tau_*^2 P \otimes D_n$. However, It may not be sampled within the desired interval $(-1, 1)$ and/or not satisfy the polynomial conditions, that is, all the roots of the polynomials are outside the unit circle. Then we will reject the sample with probability one.

Given vecB, the full conditional distribution of σ^2 follows;

$$\sigma^2|\text{vecB}, y, X_1^{p,n} \sim \mathcal{G}^{-1}(\nu_{**}/2, \lambda_{**}/2) \tag{10}$$

where $\nu_{**} = \nu_* + N$ and $\lambda_{**} = \lambda_* + e'e$.

Table 1 shows the simulation results of ARNN(1,2) using 6000 iterations and discarding the first 1000 iterations. The simulated data are generated as follows:

1. Set $N = 50$
2. Generate coordinate data from $\chi^2(8)$ and $\chi^2(6)$, respectively.
3. Generate y_1 from $\mathcal{N}(0, 0.5^2 I_N)$.
4. Generate y_t from
   $0.8y_{t-1} + 0.6W_1 y_{t-1} + 0.1W_2 y_{t-1} + u, \quad u \sim \mathcal{N}(0, 0.5^2 I_N), \quad t = 2, \cdots, 5.$

And we use the hyper-parameters as follows:

$$\tau_* = 0.01,\ \nu_* = 2,\ \lambda_* = 0.01.$$

From the table, we can find that the posterior means are estimated around true value and the MSEs are very small.

2.3 Model selection

Since we have to choose the lag and nearest neighbor order, model selection is one of the important issues in ARNN model. Familiar order selection is done by information criteria like AIC and BIC. They are calculated as follows;

$$\begin{aligned} AIC(\text{vecB}, \sigma^2) &= -2\ln(L(y|X_1^{p,n}, \text{vecB}, \sigma^2)) + 2k, \\ BIC(\text{vecB}, \sigma^2) &= -2\ln(L(y|X_1^{p,n}, \text{vecB}, \sigma^2)) + k\ln(N), \end{aligned}$$

where k is the number of parameters.

However, if we also want to compare the validity of nearest neighbor matrix, that is, we choose the distance when we use the different distances in making weight matrix, it is difficult to compare the models by AIC or BIC.

In a Bayesian framework, alternative models are usually compared by marginal likelihoods and/or by Bayes factors. Then, we calculate the marginal likelihood by Chib's (1995) method. The formula is in Appendix.

This approach can be also use to test for outliers. We simply extend the univariate ARNN model by an additive dummy variable $D_k, k = 1, ..., n$. We write the simple Bayesian ARNN(p, n) with outliers as

$$y = X_1^{p,n}\text{vecB} + D_k\gamma + u, \ \ , k = 1, \cdots, n, \ u \sim \mathcal{N}(0, \sigma^2 I_N). \tag{11}$$

and then we can test or calculate the marginal likelihoods.

3 Extension of ARNN(p, n) model

3.1 The ARXNN(p, n) model

We can extend the univariate ARXNN(p, n) model by extending the regressor matrix by another exogenous variable, which follows also a space-time pattern as the dependent variable.

$$y = X_1^{p,n}\text{vecB}_1 + X_2^{p,n}\text{vecB}_2 + u, \ u \sim \mathcal{N}(0, \sigma^2 I_N). \tag{12}$$

Now the second regressor matrix $X_2^{p,n}$ is built up in the same way from the observed exogenous $N \times T$ panel matrix X as for the first variable $X_1^{p,n}$, i.e.

$$X_2^{p,n} = (x_{t-1}, x_{t-1}^1, \cdots, x_{t-1}^n, \cdots, x_{t-p}, x_{t-p}^1, \cdots, x_{t-p}^n),$$

with $x^k_{t-j} = W_k x_{t-j}$ that is the k-th nearest neighbor of the time lag j.

This model can be easily estimated by MCMC. Let Z and vecB be $(X_1^{p,n'} X_2^{p,n'})'$ and $\mathrm{vec}(\mathrm{B}_1, \mathrm{B}_2)$, respectively and change the prior distribution as

$$\mathcal{N}(0, \tau_*^2 P \otimes D)$$

where $D = diag(D_n, D_n)$. If we replace $X_1^{p,n}$ and D_n in (9) and (10) by Z and D, we can use the same MCMC sampling methods.

Table 2 shows the simulation results of ARXNN(1,2) using 6000 iterations and discarding the first 1000 iterations. The simulated data are generated as follows:

1. Set $N = 50$
2. Generate coordinate data from $\chi^2(8)$ and $\chi^2(6)$, respectively.
3. Generate x_t from $\mathcal{N}(0, I_N)$ for $t = 1, \cdots, T$.
4. Generate y_1 from $\mathcal{N}(0, 0.5^2 I_N)$.
5. Generate y_t from

$$0.8y_{t-1} + 0.6W_1 y_{t-1} + 0.1W_2 y_{t-1} + 0.3x_{t-1} + 0.2W_1 x_{t-1} + 0.1W_2 x_{t-1} + u,$$

$$u \sim \mathcal{N}(0, 0.5^2 I_N), \quad t = 2, \cdots, 5.$$

And we use the same hyper-parameters as ARNN(p, n) model in the previous section. From the table, we can also find that the posterior means are estimated around true value and the MSEs are very small.

3.2 Hierarchical ARNN(p, n) model

Note that because the dependent variable is essentially a multivariate dynamic matrix observation we can specify the model similar to a SUR system with a hierarchical prior for the coefficients. We assume that the cross sections are correlated across time for each year, i.e.,

$$\begin{aligned} \mathrm{vecB} \sim \mathcal{N}(0, \Sigma \otimes \tau^2 D_n), \quad \sigma^2 \sim \mathcal{G}^{-1}(\nu_{\sigma *}/2, \lambda_{\sigma *}/2), \\ \tau^2 \sim \mathcal{G}^{-1}(\nu_{\tau *}/2, \lambda_{\tau *}/2), \quad \Sigma^{-1} \sim \mathcal{W}(\eta_*, S_*). \end{aligned} \tag{13}$$

Then, we can estimate the model from the following full conditional distributions:[1]

$$\mathrm{vecB}|\sigma^2,\tau^2,\Sigma,y,X_1^{p,n} \sim \mathcal{N}(\mathrm{vecB}_{**},H_{**}), \quad (14)$$

$$\sigma^2|\mathrm{vecB},\tau^2,\Sigma,y,X_1^{p,n} \sim \mathcal{G}^{-1}(\nu_{\sigma**}/2,\lambda_{\sigma**}/2), \quad (15)$$

$$\tau^2|\mathrm{vecB},\sigma^2,\Sigma,y,X_1^{p,n} \sim \mathcal{G}^{-1}(\nu_{\tau**}/2,\lambda_{\tau**}/2), \quad (16)$$

$$\Sigma^{-1}|\mathrm{vecB},\sigma^2,\tau^2,y,X_1^{p,n} \sim \mathcal{W}(\eta_{**},S_{**}), \quad (17)$$

where $\mathrm{vecB}_{**} = H(\sigma^{-2}X_1^{p,n'}y)$, $H_{**} = \{\sigma^{-2}X_1^{p,n'}X_1^{p,n} + \tau^{-2}(\Sigma\otimes D_n^{-1})\}^{-1}$, $\nu_{\sigma**} = N+\nu_{\sigma*}$, $\lambda_{\sigma**} = e'e+\lambda_{\sigma*}$, $e = y-X_1^{p,n}\mathrm{vecB}$, $\nu_{\tau**} = p(n+1)+\nu_{\tau*}$, $\lambda_{\tau**} = \mathrm{vecB}'(\Sigma\otimes D_n)^{-1}\mathrm{vecB}+\lambda_{\tau*}$, $\eta_{**} = n+1+\eta_*$ and $S_{**} = (B'D_n^{-1}B+S_*^{-1})^{-1}$.

Table 3 shows the simulation results of hierarchical ARNN(2,2) using 6000 iterations and discarding the first 1000 iterations. The simulated data are generated as follows:

1. Set $N = 50$
2. Generate coordinate data from $\chi^2(8)$ and $\chi^2(6)$, respectively.
3. Suppose $\sigma^2 = 0.05$, $\tau^2 = 0.5$ and $\Sigma = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}$.
4. Generate vecB from $\mathcal{N}(0,\Sigma\otimes\tau^2 D_n)$
5. Generate y_1 from $\mathcal{N}(0,\sigma^2 I_N)$.
6. Generate y_2 from $[y_1,W_1y_1,W_2,y_1]\beta_1+u, \quad u\sim\mathcal{N}(0,\sigma^2 I_N)$.
7. Generate y_t from $[y_{t-1},W_1y_{t-1},W_2,y_{t-1},y_{t-2},W_1y_{t-2},W_2,y_{t-2}]\mathrm{vecB} + u_t, \quad u_t\sim\mathcal{N}(0,\sigma^2 I_N)$.

And we use the following hyper-parameters.

$$\nu_{\sigma*} = 0.01,\ \lambda_{\sigma*} = 0.01,\ \nu_{\tau*} = 0.01,\ \lambda_{\tau*} = 0.01,\ \eta_* = p+1,\ S_* = S, \quad (18)$$

where S is also tightness prior, $S = diag(1,1/2,\cdots,1/p)$.

From the table, we can also find that the posterior means are estimated around true value and the MSEs are very small.

[1] The derivation of full conditional distributions are in Appendix A.

3.3 Hierarchical ARXNN(p, n) model

Next, we will consider the hierarchical ARXNN(p, n) model. We assume like the hierarchical ARNN(p, n) model that the cross sections are correlated across time for each year, i.e.,

$$\text{vecB}_1 \sim \mathcal{N}(0, \Sigma_1 \otimes \tau_1^2 D_n),\ \tau_1^2 \sim \mathcal{G}^{-1}(\nu_{\tau_1 *}/2, \lambda_{\tau_1 *}/2),\ \Sigma_1^{-1} \sim \mathcal{W}(\eta_{1*}, S_{1*}),$$
$$\text{vecB}_2 \sim \mathcal{N}(0, \Sigma_2 \otimes \tau_2^2 D_n),\ \tau_2^2 \sim \mathcal{G}^{-1}(\nu_{\tau_2 *}/2, \lambda_{\tau_2 *}/2),\ \Sigma_2^{-1} \sim \mathcal{W}(\eta_{2*}, S_{2*}),$$
$$\sigma^2 \sim \mathcal{G}^{-1}(\nu_{\sigma_2 *}/2, \lambda_{\sigma_2 *}/2).$$

Then, we can estimate the model from the following full conditional distributions:[2]

$$\begin{aligned}
\text{vecB}_i|\text{vecB}_{-i}, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n} &\sim \mathcal{N}(\text{vecB}_{i**}, H_{i**}),\\
\sigma^2|\text{vecB}_1, \text{vecB}_2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n} &\sim \mathcal{G}^{-1}(\nu_{\sigma**}/2, \lambda_{\sigma**}/2),\\
\tau_i^2, |\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_{-i}^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n} &\sim \mathcal{G}^{-1}(\nu_{\tau_i**}/2, \lambda_{\tau_i**}/2),\\
\Sigma_i^{-1}|\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_{-i}, y, X_1^{p,n}, X_2^{p,n} &\sim \mathcal{W}(\eta_{i**}, S_{i**}), \qquad (19)
\end{aligned}$$

where $\text{vecB}_i = H_{i**}(\sigma^{-2} X_i^{p,n'}(y - X_{-i}^{p,n}\text{vecB}_{-i}))$, $H_{i**} = (\sigma^{-2} X_i^{p,n'} X_i^{p,n} + \tau_i^{-2}(\Sigma_i \otimes D_n)^{-1})^{-1}$, $\nu_{\sigma**} = N + \nu_{\sigma*}$, $\lambda_{\sigma**} = e'e + \lambda_{\sigma*}$, $e = y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2$, $\nu_{\tau_i**} = n + 1 + \nu_{\tau_i*}$, $\lambda_{\tau_i**} = \text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i + \lambda_{\tau_i*}$, $\eta_{i**} = n + 1 + \eta_{i*}$ and $S_{i**} = (\text{B}_i' D_n^{-1} \text{B}_i + S_{i*}^{-1})^{-1}$.

Table 4 shows the simulation results of hierarchical ARXNN(2,2) using 6000 iterations and discarding the first 1000 iterations. The simulated data are generated as follows:

1. Set $N = 50$
2. Generate coordinate data from $\chi^2(8)$ and $\chi^2(6)$, respectively.
3. Suppose $\sigma^2 = 0.05$, $\tau_1^2 = 0.5$, $\tau_2^2 = 0.5$ and $\Sigma_1 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.4 \end{bmatrix}$ and $\Sigma_2 = \begin{bmatrix} 0.4 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}$.

[2] The derivation of full conditional distributions are also in Appendix B.

4. Generate vecB_1 and vecB_2 from $\mathcal{N}(0, \Sigma_1 \otimes \tau^2 D_n)$ and $\mathcal{N}(0, \Sigma_2 \otimes \tau^2 D_n)$, respectively.

5. Generate x_t from $\mathcal{N}(0, I_N)$ for $t = 1, \cdots, T$.

6. Generate y_1 from $\mathcal{N}(0, \sigma^2 I_N)$.

7. Generate y_2 from $[y_1, W_1 y_1, W_2, y_1]\beta_1 + [x_1, W_1 x_1, W_2, x_1]\gamma_1 + u, \quad u \sim \mathcal{N}(0, \sigma^2 I_N)$, where γ_1 is the first column of vecB_2.

8. Generate y_t from $[y_{t-1}, W_1 y_{t-1}, W_2, y_{t-1}, y_{t-2}, W_1 y_{t-2}, W_2, y_{t-2}]\mathrm{vecB}_1 + [x_{t-1}, W_1 x_{t-1}, W_2, x_{t-1}, x_{t-2}, W_1 x_{t-2}, W_2, x_{t-2}]\mathrm{vecB}_2 + u_t, \quad u_t \sim \mathcal{N}(0, \sigma^2 I_N)$.

From the table, we can also find that the posterior means are estimated around true value and the MSEs are very small.

4 Empirical results

4.1 Data set

First, we will explain the data set. We use the growth rates of Gross Domestic Product (GDP) of 227 regions in central Europe from 1995 to 2001. We use GDP in real term (1995 = 100), take log from and we use centered, i.e. de-meaned data: $GDP_{it} - G\bar{D}P$, where $G\bar{D}P = N^{-1}\sum_{i=1}^{N} GDP_{it}$. As an exogenous variable, we consider the population data. As same as the endogenous variable, we take log from and use centered, i.e. de-meaned data. To construct nearest neighbors, we need some kind of distance metric between the regions. As we mentioned in the previous section, we want to compare different type of weight matrices. First of all, we use the coordinate data of the cell centers and secondly, we use travel time data to construct the nearest neighbor matrix.

4.2 The results of the ARNN estimation

For the tightness prior distributions, the hyper-parameters are specified as follows;

$$\tau_* = 0.01,\ \nu_* = 2,\ \lambda_* = 0.01.$$

We ran the MCMC algorithm, using 6000 iterations and discarding the first 1000 iterations.

First of all, we have to choose the numbers of lags and neighbors and weight matrix. Table 5 shows the results of the AIC, BIC estimation, log marginal likelihood and the acceptance rate. From Table 5 we see that both AIC and BIC are minimal for the values $p = 4$ and $n = 1$ and $p = 1$ and $n = 1$, respectively, when we use the coordinate data. However, when we use as distance metric the travel time data, both the AIC and BIC criteria take the minimum for the values of $p = 1$ and $n = 3$. Therefore, we can not say which model is the best by AIC or BIC. When we compare the marginal likelihood of $p = 1$ and $n = 3$ with coordinate data to with travel time data, we can find that the using the model ARNN(1,3) with travel time data is the best model in ARNN. Therefore we will choose the ARNN(1,3) model based on coordinate data. Furthermore we can see that the acceptance rate becomes smaller as the numbers of p and n increases.

4.3 The results of the ARXNN estimation

For the tightness prior distributions, we use the same hyper-parameter in the previous subsection. And we ran the MCMC algorithm, using 6000 iterations and discarding the first 1000 iterations.

First of all, we also have to choose the numbers of lags and neighbors and weight matrix. Table 6 shows the results of the AIC, BIC estimation, marginal likelihood and the acceptance rate. From Table 6 we see that both AIC and BIC are minimal for the values $p = 1$ and $n = 1$, when we use the coordinate data. However, when we use as distance metric the travel time data, the AIC and BIC criteria take the minimum for the values of $p = 1$ and $n = 3$ and $p = 1$ and $n = 1$, respectively. Therefore, we can not say which model is the best in this class of model. When we compare the marginal likelihood, we can find that ARXNN(1,1) using travel time data is the best model. Therefore we will choose the ARXNN(1,1) model based on travel time data.

4.4 The results of the hierarchical ARNN estimation

For the tightness prior distributions, the hyper-parameters are specified as follows;

$$\nu_{\sigma*} = 0.01,\ \lambda_{\sigma*} = 0.01,\ \nu_{\tau*} = 0.01,\ \lambda_{\tau*} = 0.01,\ \eta_* = p + 1,\ S_* = S.$$

We ran the MCMC algorithm, using 6000 iterations and discarding the first 1000 iterations.

First of all, we also have to choose the numbers of lags and neighbors and weight matrix. Table 7 shows the results of the marginal likelihood and the acceptance rate. In hierarchical model, as we cannot evaluate by AIC or BIC, we will compare the models by marginal likelihood. From Table 7, when we compare the marginal likelihood, we can find that the the hierarchical ARNN(3,2) model with travel time data is the best model in the class of hierarchical ARNN model.

4.5 The results of the hierarchical ARXNN estimation

For the tightness prior distributions, the hyper-parameters are specified as follows;

$$\nu_{\sigma*} = 0.01,\ \lambda_{\sigma*} = 0.01,\ \nu_{\tau_1*} = 0.01,\ \lambda_{\tau_1*} = 0.01,\ \nu_{\tau_2*} = 0.01,$$
$$\lambda_{\tau_2*} = 0.01,\ \eta_{1*} = p + 1,\ S_{1*} = S,\ \eta_{2*} = p + 1,\ S_{2*} = S.$$

We ran the MCMC algorithm, using 6000 iterations and discarding the first 1000 iterations.

First of all, we also have to choose the numbers of lags and neighbors and weight matrix. Table 8 shows the results of the marginal likelihood and the acceptance rate. From Table 8, when we compare the marginal likelihood, we can find that the the hierarchical ARXNN(3,4) model with travel time data is the best model in the class of hierarchical ARNN model.

4.6 Posterior means

Table 9 shows the posterior means and standard deviations of ARNN(1,3) model. From the result, we find that the serial correlation is not significant and small. On the other hand, the spatial correlation is larger than serial correlation and NN3 is significant. It implies that the economic activity affets to the third nearest neighbors.

5 Conclusion

This paper has defined a new class of spatio-temporal models, and we estimated the autoregressive nearest neighbor (ARNN) model from a Bayesian point of view and proposed the tightness prior for the model. We derived the joint posterior distribution and proposed MCMC methods to estimate the parameters of the model and extended to the model with exogenous variables. We examined the regional GDP dynamics of 227 regions in six countries of central Europe during the period 1995 to 2001. Our results show a high spatial correlation and a rather small serial (time) correlation in the estimation of regional GDP.

Appendix A: Calculation of marginal likelihood

The calculation of marginal likelihood from the Gibbs output is shown in Chib (1995) in detail. However, we will sketch the calculation way, briefly.

Under model M_k, let $L(y|\theta_k, M_k)$ and $p(\theta_k|M_k)$ be likelihood and prior for the model, respectively. Then, the marginal likelihood of the model is defined as

$$m(y) = \int L(y|\theta_k, M_k)p(\theta_k|M_k). \qquad (20)$$

Since the marginal likelihood can be written as

$$m(y) = \frac{L(y|\theta_k, M_k)p(\theta_k|M_k)}{p(\theta_k|y, M_k)}, \qquad (21)$$

Chib (1995) suggests to estimate the marginal likelihood from the expression

$$\log m(y) = \log L(y|\theta_k^*, M_k) + \log p(\theta_k^*|M_k) - \log p(\theta_k^*|y, M_k), \qquad (22)$$

where θ_k^* is a particular high density point (typically the posterior mean or the ML estimate). He also provides a computationally efficient method to estimate the posterior ordinate $p(\theta_k^*|y, M_k)$ in the context of Gibbs sampling.

The method in our model is as follows: In ARNN model, for example, we set $\theta_k = (\text{vecB}, \sigma^2)$ and estimate the posterior ordinate $p(\theta_k^*|y, M_k)$ via the decomposition

$$p(\theta_k^*|y, M_k) = p(\text{vecB}^*|\sigma^{*2}, y)p(\sigma^{*2}|\text{vecB}^*, y). \tag{23}$$

$p(\text{vecB}^*|\sigma^{*2}, y)$ and $p(\sigma^{*2}|\text{vecB}^*, y)$ are calculated from the Gibbs output as follows:

$$\begin{aligned} p(\text{vecB}^*|\sigma^{*2}, y) &= \frac{1}{iter}\sum_{g=1}^{iter} p(\text{vecB}^*|\text{vecB}_{**}^{(g)}, \Sigma_{**}^{(g)}), & (24) \\ p(\sigma^{*2}|\text{vecB}^*, y) &= \frac{1}{iter}\sum_{g=1}^{iter} p(\sigma^{*2}|\nu_{**}/2, \lambda_{**}^{(g)}/2), & (25) \end{aligned}$$

where, it should be noted, $\text{vecB}_{**}^{(g)}$, $\Sigma_{**}^{(g)}$ and $\lambda_{**}^{(g)}$ are produced as a by-product of the sampling.

Appendix B: Hierarchical ARNN(p, n) model

Posterior distribution of hierarchical ARNN (p, n) model is written as

$$\begin{aligned} p(\text{vecB}, \sigma^2, \Sigma, \tau^2|y, X_1^{p,n}) &\propto L(y|\text{vecB}, \sigma^2, X_1^{p,n})p(\text{vecB}, \sigma^2, \tau^2, \Sigma), \\ &\propto L(y|\text{vecB}, \sigma^2 X_1^{p,n})p(\text{vecB}|\tau^2, \Sigma)p(\sigma^2)p(\tau^2)p(\Sigma), \\ &\propto (\sigma^2)^{-\frac{N}{2}} \exp\left\{ -\frac{(y - X_1^{p,n}\text{vecB})'(y - X_1^{p,n}\text{vecB})}{2\sigma^2} \right\} \\ &\quad \times|\Sigma \otimes \tau^2 D_n|^{-\frac{1}{2}} \exp\left\{ -\frac{\text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB}}{2\tau^2} \right\} \\ &\quad \times(\sigma^2)^{-(\frac{\nu_{\sigma*}}{2}+1)} \exp\left\{ -\frac{\lambda_{\sigma*}}{2\sigma^2} \right\} \\ &\quad \times(\tau^2)^{-(\frac{\nu_{\tau*}}{2}+1)} \exp\left\{ -\frac{\lambda_{\tau*}}{2\tau^2} \right\} \\ &\quad \times|\Sigma^{-1}|^{\frac{\eta_*-p-1}{2}} \exp\left\{ -\frac{1}{2}tr(\Sigma^{-1}S_*^{-1}) \right\}. \end{aligned} \tag{26}$$

Then, the full conditional distribution of vecB is as follows:

$$\begin{aligned} p(\text{vecB}|\sigma^2, \tau^2, \Sigma, y, X_1^{p,n}) &\propto \exp\left\{ -\frac{(y - X_1^{p,n}\text{vecB})'(y - X_1^{p,n}\text{vecB})}{2\sigma^2} \right\} \\ &\quad \times \exp\left\{ -\frac{\text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB}}{2\tau^2} \right\}, \\ &\propto \mathcal{N}(\text{vecB}_{**}, H_{**}), \end{aligned} \tag{27}$$

where $\text{vecB}_{**} = H_{**}(\sigma^{-2} X_1^{p,n'} y)$ and $H_{**} = \{\sigma^{-2} X_1^{p,n'} X_1^{p,n} + \tau^{-2} (\Sigma \otimes D_n)^{-1}\}^{-1}$.

The full conditional distribution of σ^2 is as follows:

$$
\begin{aligned}
p(\sigma^2 | \text{vecB}, \tau^2, \Sigma, y, X_1^{p,n}) &\propto (\sigma^2)^{-\frac{N}{2}} \exp\left\{ - \frac{(y - X_1^{p,n}\text{vecB})'(y - X_1^{p,n}\text{vecB})}{2\sigma^2} \right\} \\
&\quad \times (\sigma^2)^{-(\frac{\nu_{\sigma *}}{2}+1)} \exp\left\{ - \frac{\lambda_{\sigma *}}{2\sigma^2} \right\}, \\
&\propto \mathcal{G}^{-1}(\nu_{\sigma **}/2, \lambda_{\sigma **}/2), \qquad (28)
\end{aligned}
$$

where $\nu_{\sigma **} = N + \nu_{\sigma *}$, $\lambda_{\sigma **} = e'e + \lambda_{\sigma *}$ and $e = y - X_1^{p,n}\text{vecB}$.

The full conditional distribution of τ^2 is as follows:

$$
\begin{aligned}
p(\tau^2 | \text{vecB}, \sigma^2, \Sigma, y, X_1^{p,n}) &\propto |\Sigma \otimes \tau^2 D_n|^{-\frac{1}{2}} \exp\left\{ - \frac{\text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB}}{2\tau^2} \right\} \\
&\quad \times (\tau^2)^{-(\frac{\nu_{\tau *}}{2}+1)} \exp\left\{ - \frac{\lambda_{\tau *}}{2\tau^2} \right\} \\
&\propto (\tau^2)^{-\frac{p(n+1)}{2}} \exp\left\{ - \frac{\text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB}}{2\tau^2} \right\} \\
&\quad \times (\tau^2)^{-(\frac{\nu_{\tau *}}{2}+1)} \exp\left\{ - \frac{\lambda_{\tau *}}{2\tau^2} \right\} \\
&\propto \mathcal{G}^{-1}(\nu_{\tau **}/2, \lambda_{\tau **}/2), \qquad (29)
\end{aligned}
$$

where $\nu_{\tau **} = p(n+1) + \nu_{\tau *}$ and $\lambda_{\tau **} = \text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB} + \lambda_{\tau *}$

Finally, the full conditional distribution of Σ is as follows:

$$
\begin{aligned}
p(\Sigma^{-1} | \text{vecB}, \sigma^2, \tau^2, y, X_1^{p,n}) &\propto |\Sigma \otimes \tau^2 D_n|^{-\frac{1}{2}} \exp\left\{ - \frac{\text{vecB}'(\Sigma \otimes D_n)^{-1}\text{vecB}}{2\tau^2} \right\} \\
&\quad \times |\Sigma^{-1}|^{\frac{\eta_* - p - 1}{2}} \exp\left\{ - \frac{1}{2} tr(\Sigma^{-1} S_*^{-1}) \right\} \\
&\propto |\Sigma^{-1}|^{\frac{n+1}{2}} \exp\left\{ - \frac{1}{2} tr(\Sigma^{-1} B' D_n^{-1} B) \right\} \\
&\quad \times |\Sigma^{-1}|^{\frac{\eta_* - p - 1}{2}} \exp\left\{ - \frac{1}{2} tr(\Sigma^{-1} S_*^{-1}) \right\} \\
&\propto \mathcal{W}(\eta_{**}, S_{**}), \qquad (30)
\end{aligned}
$$

where $\eta_{**} = n + 1 + \eta_*$ and $S_{**} = (B' D_n^{-1} B + S_*^{-1})^{-1}$.

Appendix C: Hierarchical ARXNN(p, n) model

Posterior distribution of hierarchical ARXNN (p, n) model is written as

$p(\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2 | y, X_1^{p,n}, X_2^{p,n})$

$$
\begin{aligned}
&\propto \quad L(y|\text{vecB}_1, \text{vecB}_2, \sigma^2 y, X_1^{p,n}, X_2^{p,n}) p(\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2), \\
&\propto \quad L(y|\text{vecB}, \sigma^2) p(\text{vecB}_1|\tau_1^2, \Sigma_1) p(\text{vecB}_2|\tau_2^2, \Sigma_2) p(\sigma^2) p(\tau_1^2) p(\tau_2^2) p(\Sigma_1) p(\Sigma_2), \\
&\propto \quad (\sigma^2)^{-\frac{N}{2}} \exp\Big\{ -\frac{(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)'(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)}{2\sigma^2} \Big\} \\
&\qquad \times |\Sigma_1 \otimes \tau_1^2 D_n|^{-\frac{1}{2}} \exp\Big\{ -\frac{\text{vecB}_1'(\Sigma_1 \otimes D_n)^{-1}\text{vecB}_1}{2\tau_1^2} \Big\} \\
&\qquad \times |\Sigma_2 \otimes \tau_2^2 D_n|^{-\frac{1}{2}} \exp\Big\{ -\frac{\text{vecB}_2'(\Sigma_2 \otimes D_n)^{-1}\text{vecB}_2}{2\tau_2^2} \Big\} \\
&\qquad \times (\sigma^2)^{-(\frac{\nu_{\sigma *}}{2}+1)} \exp\Big\{ -\frac{\lambda_{\sigma *}}{2\sigma^2} \Big\} \\
&\qquad \times (\tau_1^2)^{-(\frac{\nu_{\tau_1 *}}{2}+1)} \exp\Big\{ -\frac{\lambda_{\tau_1 *}}{2\tau_1^2} \Big\} \\
&\qquad \times (\tau_2^2)^{-(\frac{\nu_{\tau_2 *}}{2}+1)} \exp\Big\{ -\frac{\lambda_{\tau_2 *}}{2\tau_2^2} \Big\} \\
&\qquad \times |\Sigma_1^{-1}|^{\frac{\eta_{1*}-p-1}{2}} \exp\Big\{ -\frac{1}{2} tr(\Sigma_1^{-1} S_{1*}^{-1}) \Big\} \\
&\qquad \times |\Sigma_2^{-1}|^{\frac{\eta_{2*}-p-1}{2}} \exp\Big\{ -\frac{1}{2} tr(\Sigma_2^{-1} S_{2*}^{-1}) \Big\}.
\end{aligned} \tag{31}
$$

Then, the full conditional distribution of vecB_i for $i = 1, 2$ is as follows:

$$
\begin{aligned}
&p(\text{vecB}_i|\text{vecB}_{-i}, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n}) \\
&\qquad \propto \quad \exp\Big\{ -\frac{(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)'(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)}{2\sigma^2} \Big\} \\
&\qquad\qquad \times |\Sigma_i \otimes \tau_i^2 D_n|^{-\frac{1}{2}} \exp\Big\{ -\frac{\text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i}{2\tau_i^2} \Big\} \\
&\qquad \propto \quad \mathcal{N}(\text{vecB}_{i**}, H_{i**}),
\end{aligned} \tag{32}
$$

where $\text{vecB}_i = H_{i**}(\sigma^{-2} X_i^{p,n'}(y - X_{-i}^{p,n}\text{vecB}_{-i}))$ and $H_{i**} = (\sigma^{-2} X_i^{p,n'} X_i^{p,n} + \tau_i^{-2}(\Sigma_i \otimes D_n)^{-1})^{-1}$.

The full conditional distribution of σ^2 is as follows:

$$
\begin{aligned}
&p(\sigma^2|\text{vecB}_1, \text{vecB}_2, \tau_1^2, \tau_2^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n}) \\
&\qquad \propto \quad (\sigma^2)^{-\frac{N}{2}} \exp\Big\{ -\frac{(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)'(y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2)}{2\sigma^2} \Big\} \\
&\qquad\qquad \times (\sigma^2)^{-(\frac{\nu_{\sigma *}}{2}+1)} \exp\Big\{ -\frac{\lambda_{\sigma *}}{2\sigma^2} \Big\} \\
&\qquad \propto \quad \mathcal{G}^{-1}(\nu_{\sigma **}/2, \lambda_{\sigma **}/2),
\end{aligned} \tag{33}
$$

where $\nu_{\sigma **} = N + \nu_{\sigma *}$, $\lambda_{\sigma **} = e'e + \lambda_{\sigma *}$ and $e = y - X_1^{p,n}\text{vecB}_1 - X_2^{p,n}\text{vecB}_2$.

Then, the full conditional distribution of τ_i^2 for $i = 1, 2$ is as follows:

$$\begin{aligned}
& p(\tau_i^2, |\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_{-i}^2, \Sigma_1, \Sigma_2, y, X_1^{p,n}, X_2^{p,n}) \\
& \quad \propto \quad |\Sigma_i \otimes \tau_i^2 D_n|^{-\frac{1}{2}} \exp\bigg\{ - \frac{\text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i}{2\tau_i^2} \bigg\} \\
& \qquad \quad \times (\tau_i^2)^{-(\frac{\nu_{\tau_i *}}{2}+1)} \exp\bigg\{ - \frac{\lambda_{\tau_i *}}{2\tau_i^2} \bigg\} \\
& \quad \propto \quad \tau_i^{-\frac{n+1}{2}} \exp\bigg\{ - \frac{\text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i}{2\tau_i^2} \bigg\} \\
& \qquad \quad \times (\tau_i^2)^{-(\frac{\nu_{\tau_i *}}{2}+1)} \exp\bigg\{ - \frac{\lambda_{\tau_i *}}{2\tau_i^2} \bigg\} \\
& \quad \propto \quad \mathcal{G}^{-1}(\nu_{\tau_i **}/2, \lambda_{\tau_i **}/2), \qquad (34)
\end{aligned}$$

where $\nu_{\tau_i **} = n + 1 + \nu_{\tau_i *}$ and $\lambda_{\tau_i **} = \text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i + \lambda_{\tau_i *}$.

Finally, the full conditional distribution of Σ_i for $i = 1, 2$ is as follows:

$$\begin{aligned}
& p(\Sigma_i^{-1}|\text{vecB}_1, \text{vecB}_2, \sigma^2, \tau_1^2, \tau_2^2, \Sigma_{-i}, y, X_1^{p,n}, X_2^{p,n}) \\
& \quad \propto \quad |\Sigma_i \otimes \tau_i^2 D_n|^{-\frac{1}{2}} \exp\bigg\{ - \frac{\text{vecB}_i'(\Sigma_i \otimes D_n)^{-1}\text{vecB}_i}{2\tau_i^2} \bigg\} \\
& \qquad \quad \times |\Sigma_i^{-1}|^{\frac{\eta_{i*}-p-1}{2}} \exp\bigg\{ - \frac{1}{2} tr(\Sigma_i^{-1} S_{i*}^{-1}) \bigg\} \\
& \quad \propto \quad |\Sigma_i|^{-\frac{n+1}{2}} \exp\bigg\{ - \frac{1}{2\tau_i^2} tr(\Sigma_i^{-1} \text{B}_i' D_n^{-1} \text{B}_i) \bigg\} \\
& \qquad \quad \times |\Sigma_i^{-1}|^{\frac{\eta_{i*}-p-1}{2}} \exp\bigg\{ - \frac{1}{2} tr(\Sigma_i^{-1} S_{i*}^{-1}) \bigg\} \\
& \quad \propto \quad \mathcal{W}(\eta_{i**}, S_{i**}), \qquad (35)
\end{aligned}$$

where $\eta_{i**} = n + 1 + \eta_{i*}$ and $S_{i**} = (\text{B}_i' D_n^{-1} \text{B}_i + S_{i*}^{-1})^{-1}$.

References

[1] Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer.

[2] Banerjee, S., B.P. Carlin and A.E. Gelfand (2003) *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall.

[3] Chib, S. (1995) "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, **90**, 1313–1321.

Table 1: Simulation result of ARNN(1,2): Posterior means, standard deviations (in parenthes) and MSE

	True value	Estimated	MSE
AR(1)	0.800	0.797	0.009
		(0.092)	
NN(1)	0.600	0.631	0.014
		(0.116)	
NN(2)	0.100	0.113	0.016
		(0.126)	
σ^2	0.500	0.693	0.058
		(0.145)	

[4] Gelfand, A.E. and A.F.M. Smith (1990) "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, **85**, 398–409.

[5] Kakamu, K. and H. Wago (2005) "Bayesian Panel Spatial Autoregressive Probit Model with an Application to Business Cycle in Japan," *mimeo*.

Table 2: Simulation result of ARXNN(1,2): Posterior means, standard deviations (in parenthes) and MSE

	True value	Estimated	MSE
AR(1)	0.800	0.682	0.024
		(0.099)	
NN(1)	0.600	0.739	0.036
		(0.129)	
NN(2)	0.100	0.113	0.012
		(0.107)	
XAR(1)	0.300	0.308	0.006
		(0.080)	
XNN(1)	0.200	0.416	0.058
		(0.108)	
XNN(2)	0.100	-0.035	0.044
		(0.160)	
σ^2	0.500	0.433	0.013
		(0.093)	

Table 3: Simulation result of hierarchical ARNN(2,2): Posterior means, standard deviations (in parenthes) and MSE

	True value	Estimated	MSE
AR1	0.061	0.061	0.013
		(0.115)	
NN(1)	-0.177	0.056	0.076
		(0.145)	
NN(2)	0.372	0.249	0.058
		(0.208)	
AR2	0.489	0.488	0.014
		(0.117)	
NN(1)	-0.391	-0.171	0.072
		(0.153)	
NN(2)	0.368	0.112	0.121
		(0.236)	
σ^2	0.050	0.041	0.000
		(0.007)	
τ^2	0.500	1.202	0.855
		(0.601)	
	True value		
	0.500	0.200	
	0.200	0.400	
	Estimated		
	0.409	0.019	
	0.019	0.807	

Table 4: Simulation result of hierarchical ARNN(2,2): Posterior means, standard deviations (in parenthes) and MSE

	True value	Estimated	MSE		True value	Estimated	MSE
AR1	0.327	0.269	0.019	XAR1	0.422	0.467	0.004
		(0.127)				(0.049)	
NN(1)	0.076	−0.048	0.028	XNN(1)	0.653	0.623	0.007
		(0.111)				(0.076)	
NN(2)	−0.286	−0.139	0.037	XNN(2)	0.016	0.030	0.010
		(0.125)				(0.100)	
AR2	0.147	0.179	0.010	XAR2	0.211	0.194	0.007
		(0.097)				(0.080)	
NN(1)	−0.015	0.040	0.011	XNN(1)	0.293	0.340	0.012
		(0.088)				(0.097)	
NN(2)	0.334	0.238	0.019	XNN(2)	0.288	0.326	0.011
		(0.100)				(0.099)	
τ_1^2	0.500	1.205	0.878	τ_2^2	0.500	1.201	0.890
		(0.618)				(0.631)	
σ^2	0.050	0.074	0.001	σ^2			
		(0.013)					
	Σ_1				Σ_2		
	True value				True value		
	0.500	0.200			0.400	0.200	
	0.200	0.400			0.200	0.300	
	Estimated				Estimeted		
	0.394	−0.015			0.562	0.116	
	−0.015	0.716			0.116	0.857	

Table 5: Information criteria, marginal likelihood and acceptance rate of ARNN model

Distance					
n	p	AIC	BIC	log marginal	acceptance
1	1	-983.282	-973.007*	483.884*	1.000
1	2	-981.478	-964.353	480.812	1.000
1	3	-982.957	-958.982	479.853	0.999
1	4	-985.098*	-954.273	479.298	0.999
1	5	-982.756	-945.081	476.577	1.000
2	1	-982.801	-969.102	483.298	1.000
2	2	-979.153	-955.178	479.082	0.999
2	3	-980.105	-945.855	477.951	0.999
2	4	-981.487	-936.963	477.385	0.999
2	5	-977.401	-922.602	474.404	0.999
3	1	-982.634	-965.509	483.290	0.994
3	2	-977.811	-946.986	478.757	0.990
3	3	-977.052	-932.528	477.332	0.992
3	4	-978.127	-919.903	476.940	0.966
3	5	-974.992	-903.068	474.637	0.967
Travel time					
n	p	AIC	BIC	log marginal	acceptance
1	1	-983.587	-973.312	484.178	1.000
1	2	-981.473	-964.348	480.953	1.000
1	3	-979.029	-955.054	478.239	0.999
1	4	-975.583	-944.759	475.357	0.999
1	5	-973.190	-935.516	472.725	1.000
2	1	-986.741	-973.041	485.256	0.991
2	2	-983.220	-959.245	481.208	0.998
2	3	-983.758	-949.509	479.828	0.997
2	4	-979.252	-934.727	476.914	0.998
2	5	-974.037	-919.238	473.639	0.995
3	1	-992.228*	-975.103*	487.698*	0.945
3	2	-985.071	-954.247	482.422	0.940
3	3	-984.264	-939.740	481.164	0.959
3	4	-985.198	-926.974	480.513	0.939
3	5	-978.328	-906.404	476.929	0.934

Table 6: Information criteria, marginal likelihood and acceptance rate of ARXNN model

Distance					
n	p	AIC	BIC	log marginal	acceptance
1	1	-979.813*	-962.688*	480.573*	1.000
1	2	-977.009	-946.184	476.076	0.999
1	3	-976.139	-931.615	473.369	1.000
1	4	-974.514	-916.290	470.553	0.999
1	5	-969.193	-897.269	466.663	0.999
2	1	-977.538	-953.563	479.065	0.999
2	2	-971.813	-927.289	473.122	0.999
2	3	-969.424	-904.350	469.679	0.999
2	4	-965.585	-879.961	466.032	0.999
2	5	-956.251	-850.077	460.965	0.999
3	1	-975.695	-944.870	478.469	1.000
3	2	-967.569	-909.345	471.671	1.000
3	3	-961.022	-875.398	467.190	0.999
3	4	-955.043	-842.020	463.246	0.999
3	5	-945.377	-804.954	458.445	0.999
Travel time					
n	p	AIC	BIC	log marginal	acceptance
1	1	-988.260	-971.135*	485.086*	1.000
1	2	-983.461	-952.636	479.803	0.999
1	3	-976.823	-932.298	474.582	1.000
1	4	-970.202	-911.978	469.858	0.999
1	5	-964.357	-892.433	465.814	0.999
2	1	-988.191	-964.217	484.470	0.999
2	2	-981.514	-936.990	478.340	0.999
2	3	-979.574	-914.500	474.701	0.999
2	4	-969.580	-883.956	468.702	0.999
2	5	-958.648	-852.474	463.207	0.999
3	1	-988.650*	-957.826	484.913	1.000
3	2	-979.076	-920.852	477.832	0.999
3	3	-975.893	-890.269	474.295	0.999
3	4	-972.646	-859.623	471.066	0.999
3	5	-957.881	-817.458	464.510	0.999

Table 7: Marginal likelihood and acceptance rate of hierarchical ARNN model

		Distance		Travel time	
n	p	log marginal	acceptance	log marginal	acceptance
1	2	469.643*	1.000	469.784	1.000
1	3	469.201	1.000	467.589	1.000
1	4	469.414	0.999	465.470	0.999
1	5	467.887	0.999	464.038	0.999
2	2	468.352	1.000	470.315	0.997
2	3	467.325	0.999	469.164	0.999
2	4	466.761	0.999	466.221	0.999
2	5	463.801	0.999	462.952	0.995
3	2	468.226	0.999	471.355*	0.980
3	3	466.216	0.997	469.761	0.981
3	4	465.720	0.925	469.263	0.927
3	5	463.029	0.942	464.795	0.881
4	2	468.186	0.898	470.870	0.934
4	3	465.845	0.760	468.785	0.877
4	4	465.804	0.605	467.858	0.641
4	5	461.961	0.614	462.642	0.599

Table 8: Marginal likelihood and acceptance rate of hierarchical ARXNN model

		Distance		Travel time	
n	p	log marginal	acceptance	log marginal	acceptance
1	2	464.105	1.000	464.003	1.000
1	3	464.013	0.999	461.824	0.999
1	4	464.485	0.999	460.160	0.999
1	5	465.588*	0.999	462.509	0.999
2	2	463.133	1.000	465.519	0.994
2	3	462.038	0.999	465.545	0.997
2	4	462.016	0.999	462.945	0.995
2	5	461.472	0.999	462.972	0.985
3	2	463.135	0.997	466.643	0.971
3	3	461.812	0.996	467.936	0.970
3	4	461.251	0.947	468.885*	0.907
3	5	460.214	0.965	466.867	0.853
4	2	463.589	0.897	466.759	0.911
4	3	461.922	0.789	467.710	0.857
4	4	461.041	0.648	468.510	0.595
4	5	461.519	0.629	465.823	0.545

Table 9: Empirical result of ARNN model with travel time data: Posterior means and standard deviations (in parenthes)

	ARNN(1,3)
AR1	0.02189
	0.06656
NN1	-0.12883
	0.11309
NN2	-0.07791
	0.18462
NN3	0.43697
	0.15925
σ^2	0.00076
	0.00007