Participatory Approach to Community Based Water Supply System

Kenshiro OGI* Ismu Rini Dwi ARI Hayeong JEONG
Kakuya MATSUSHIMA Kiyoshi KOBAYASHI

March 9, 2010

Abstract

Water scarcity due to climate change as well as inappropriate water governance is one of the important topics in the world, particularly in developing countries. Most people who live close to the water resource are not always economically-advantaged. Moreover, it might be seems that people living close to "economically-attractive water resource" are classified into lower category according to the standard of living in the country. Community based water supply system is one of strong alternatives to existing water supply system by the public sector. The community based water supply system works more effectively if it is based upon strong community network in the region.

This study conducts an empirical research on community based water supply system in Indonesian rural area. In this paper, we propose a discrete-choice model which describes the mechanism of resident's spontaneous collaboration to access water. We formulate a hypothesis that households with better community tie have ability to organize "community based" management system. In order to test it, we formulate a spatial probit model which can consider the effect of social interaction upon their choices in water supply system. Traditionally, spatial models are estimated with maximum likelihood method, however, in this paper, we adopt Markov chain Monte Carlo (MCMC) method to estimate parameters due to the difficulty in estimation of discrete-choice model with spatial interaction term. Using dataset from a field survey in Indonesia which we conducted in 2008 the spatial probit model is empirically tested to show that social interaction in the community plays an important role on resident's spontaneous collaboration to manage community-based water supply system.

*Undergraduate Student, Department of Global Engineering, Faculty of Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8540, Japan. Email: k.ogi@ky4.ecs.kyoto-u.ac.jp

1 Introduction

Water is one of the most important matters among basic needs for people to survive. It goes to say that water is an essential need, one can live for one month without food, but he/she can only live for 5 to 7 days without water.

Lack of access to a clean water supply is today's world problem. Regarding to NARBO report (2007), around 600 million people in the Asia - Pacific region live in the area with no connection to water services. In case of Indonesia, in the midst of total population 231.6 million inhabitants (2007), around 100 million people or 43% of the total population have no access to water supply. Indonesia local water company (named as PDAM) supplies around 39.7% of the citizen. And the rest 10% of them are starting to engage in a participatory approach to community based water supply system (named as HIPPAM) (Statistic Centre Board, 2005).

Singosari district consists of 17 villages covering 140.245 inhabitants, the highest populated district in Malang regency, has several natural springs with water flow more than 250 liter per second (3rd to 2nd magnitude level). But, the number of population with water connection is only 28% from total inhabitants. It indicates that most people who live near the water resource which is a subject of squeezing are not always economically-advantaged. There is a mechanism that people living near "economically-attractive water resource" cannot develop the water resource with their value.

In this sense, situation above seem in line with the Asian Water Development Outlook (AWDO) report that the future water crisis in Asian countries, it will not be because of physical scarcity of water, but because of inadequate or inappropriate water governance, including management practices, institutional arrangements, and socio-political conditions, which leave much to be desired (AWDO, 2007).

Moreover, under the situation of lack of access to water, collaboration activities for water supply system by community members who lives in near the water resource has examined in many countries. Collective action may be defined as action on the part of one or more people striving to achieve objective or satisfy common interest of the group, implies devising frameworks that limit the pursuit of individual self interest and sustain the benefit shared by the group. M. Olson in a theorem of "The Logic of Collective Action" mentions that a group cannot base its reasoning on the rational choices of its individuals, unless the number of individuals in a group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest.

In order to support and sustain the participatory approach to community based water supply system, it is necessary to clarify the mechanism and to invent institutional system for support the collaboration activities. However, there are not enough investigations on the participatory approach to community based water supply system. Necessitate of better understanding about local community structure and network is urgent toward encouraging suitable water policy and institutional restructuring.

Therefore this study is questioning: Why people are involved in establishing the community based water supply system in order to access water?, Whether their choices to join the community based water supply system are cooperated or not?, Why people do or do not engage in a kind of collective action?, It is important to investigate the mechanism of the spontaneous collaboration to access water. In other words, this study aims to investigate interdependent preference in a choice of clean water of the community on a field survey of Indonesian's water supply system.

Then the paper is structured as follows. Section 2 explains a spatial probit model and Markov Chain Monte Carlo (MCMC) for estimation method and section 3 describes the example of the empirical application. Section 4 presents concluding remarks.

2 Model and Estimation Method

2.1 Model

This section focuses on the discrete-choice model for whether an household joins HIPPAM conditional on that household's characteristics. We start by introducing the main assumptions in the model and the notation that will be used for the rest of the paper. Let n be the number of individual housholds. Each household has two alternatives, labeled as 1 for joining HIPPAM and 0 for otherwise. For each household we observe whether the household joins HIPPAM or not and model it as the realization of a random variable y_i. Economic theory suggests that the dicision to join is primarily made to maximize the discounted value of future profits, so we assume that the choice of whether to join HIPPAM or not is the result of an household's decision to maximize their utility. An event will occur with a certain probability p if the utility derived from choosing that alternatives is greater than the utility from the other alternative. Let z_i be the difference in the utility from alternatives 1 and 0. The difference in utility is modeled as:

$$z_i = {x_i}'\beta + \theta_i + \varepsilon_i \tag{1}$$

where $i = 1, \cdots, n$, $x_i = (x_{ik} : k = 1, \cdots, K)'$ is a vector of observed household specific attributes, $\beta = (\beta_k : k = 1, \cdots, K)'$ is a vector of unobserved parameters to be estimated, θ_i is an unobserved random effect component, and ε_i is the stochastic error term with $\varepsilon_i \sim \mathcal{N}(0, 1)$. We do not observe z_i, but only observe the sign of z_i. We observe the household choice y_i being equal to 1 or 0, depending on wheter z_i has a positive sign indicating the higher utility from this alternative or a negative sign associated with the lower utility associated with this alternative. Therefore we observe:

$$y_i = \begin{cases} 1 & if \quad z_i > 0 \\ 0 & if \quad z_i \leq 0 \end{cases} \tag{2}$$

The probability of choosing alternative 1 is given by:

$$P_i = P(y_i = 1) = P(z_i > 0) \tag{3}$$

The distinction between this model and the standard probit model is the term θ_i. The unobserved component θ_i is constructed such that it allows for spatial interaction among households. This is obtained by specifying θ_i according to a spatial autoregressive structure:

$$\theta_i = \rho \sum_{j=1}^{n} w_{ij}\theta_j + u_i \tag{4}$$

with $u_i \sim \mathcal{N}(0, \sigma^2)$, $W = (w_{ij} : i, j = 1, \cdots, n)$ is a row standardized spatial weight matrix such that $\sum_{j=1}^{n} w_{ij} = 1$. ρ can be interpreted as the degree of spatial dependence across households. Positive (negative) value of ρ indicates possitive (negative) correlation among households. We can write equation (4) in matrix notation:

$$\theta = \rho W \theta + u \tag{5}$$

where $u \sim \mathcal{N}_n(0_n, \sigma^2 I_n)$ and I_n is the identity matrix. Letting $S = I_n - \rho W$, we can obtain a solution for θ using (5):

$$\theta = S^{-1} u \tag{6}$$

It is worth noting that in our model, there is a network propagation effect captured in equation (5), where in the exogenous model the effect associated with a covariate does not propagate among households. Our model presents a simple test on the existence of propagation effect. If ρ is significantly different from zero, then we conclude that there could be some spatial correlation beyond what is captured in the $x_i\beta$ term in the equation (1).

From (6) we see that the distribution for θ is given by:

$$\theta | (\rho, \sigma^2) \sim \mathcal{N}_n(0_n, \sigma^2 (S'S)^{-1}) \tag{7}$$

The error term ε is assumed to be conditionally independent of the spatial unobserved component such that $\varepsilon | \theta \sim \mathcal{N}_n(0_n, {\sigma_\varepsilon}^2 I_n)$ and we assume ${\sigma_\varepsilon}^2 = 1$.

The full model in matrix notation is given by:

$$z = X\beta + \theta + \varepsilon \tag{8}$$

The likelihood function of this model as follows:

$$L(y|\beta, \theta, \rho, \sigma^2) = \prod_{i=1}^{n} \Phi\left({x_i}'\beta + (S^{-1})_i u\right)^{y_i} \left\{1 - \Phi({x_i}'\beta + (S^{-1})_i u)\right\}^{1-y_i} \tag{9}$$

where Φ and $(S^{-1})_i$ denote the cumulative distribution function of the standard normal and ith row of S^{-1}. But it is difficult to estimate by maximum likelihood since this model has complicated form. Then, we use the Bayesian inference approach to estimate each parameters of equation by using the Markov Chain monte Carlo method that sample sequentially from the complete set of conditional posterior distributions for the parameters. The MCMC methods provides a powerful tool for simulating complicated posterior distributions.

2.2 Bayesian Inference

We examine to estimate above spatial probit model with using Markov Chain Monte Carlo (MCMC) method. The Gibbs sampler was the first MCMC algorithm and was used in statistics and econometrics popularly, which arrives at the target distribution of the unknown parameters by sequentially sampling from a set of conditional distributions of the parameters. This is very useful since usually it is difficult to find an analytical result for the posterior densities. The MCMC method provides a sample from the posterior density and we can use this sample to draw inferences about the parameters of interest. Under mild regularity conditions satisfied in this application, these samples converge to sample from the posterior distribution.

Most of the parameter can be sampled with using Gibbs sampler, However, only sampling the spatial parameter ρ is difficult with using formal probability distribution. Therefore, we apply the Metropolis-Hastings (MH) sampling method.

To obtain the posterior distribution, we use the Bayes theorem and examine as follows:

$$p(\beta, \theta, \rho, \sigma^2, z|y) \propto L(y|\beta, \theta, \rho, \sigma^2, z) \cdot \pi(\beta, \theta, \rho, \sigma^2, z) \tag{10}$$

where $p(\cdot)$ represent the posterior distribution. The prior distribution of parameters β, ρ, σ^2 are assumed independent. Therefore the posterior joint density is proportionaly as the following formula:

$$p(\beta, \theta, \rho, \sigma^2, z|y) \propto L(y|z) \cdot \pi(z|\beta, \theta) \cdot \pi(\theta|\rho, \sigma^2) \cdot \pi(\beta) \cdot \pi(\rho) \cdot \pi(\sigma^2) \tag{11}$$

Using above posterior joint density, we obtain the appropriate conditional posterior distribution and examine MCMC sampling methods in the following section. Before we examine the Bayesian estimation we set each parameter's prior distributions as follows:

$$\begin{aligned} \pi(\beta) \sim \mathcal{N}_K(c, T), \quad \pi(\sigma^2) \sim \mathcal{IG}(\alpha, \nu), \quad \pi(\rho) \sim \mathcal{U}(\lambda_{min}^{-1}, \lambda_{max}^{-1}) \\ \pi(\theta|\rho, \sigma^2) \sim \mathcal{N}_n(0_n, \sigma^2(S'S)^{-1}), \quad \pi(z|\beta, \theta) \sim \mathcal{N}_n(X\beta + \theta, I_n) \end{aligned} \tag{12}$$

where β has normal conjugate prior distribution with means set to zero and covariance matrix set to $100I_K$, and σ^2 is assigned a conjugate inverted gamma prior with $\alpha = 25$ and $\nu = 3$. We employ a uniform prior distribution on ρ over a specified range. The parameter ρ must lie in the interval $[\lambda_{min}^{-1}, \lambda_{max}^{-1}]$, where λ_{min} and λ_{max} denote the minimum and maximum eigenvalues of W, for the matrix $S = I_n - \rho W$ to be invertable(Sun, Tsukawa and Speckman 1999).

Introducing each prior distribution (12) into equation (11), we can derive each parameter's posterior distribution. In the next section, we examine the MCMC sampling method with using these posterior distributions.

2.3 The Markov Chain Monte Carlo (MCMC) Sampler

We prepare the starting values for each parameters $\beta^0, \theta^0, \sigma^{2(0)}, \rho^0$ and z^0 which we designate with the superscript 0, and examine the MCMC sampling method as follows.

1. Calculate $p(\beta|\rho^0, \theta^0, \sigma^{2(0)}, z^0, y)$ using each initial parameter. We carry out a multivariate random draw to determine β^1.

$$\beta|(\theta, \rho, \sigma^2, z, y) \sim \mathcal{N}_K(A^{-1}b, A^{-1}) \tag{13}$$

   where $A = X'X + T^{-1}, b = X'(z - \theta) + T^{-1}c$.

2. Calculate $p(\theta|\beta^1, \rho^0, \sigma^{2(0)}, z^0, y)$, we carry out a multivariate random draw to determine θ^1.

$$\theta|(\beta, \rho, \sigma^2, z, y) \sim \mathcal{N}_n(A_0^{-1}b_0, A_0^{-1}) \tag{14}$$

   where $A_0 = \sigma^{-2}S'S + I_n, b_0 = z - X\beta$

3. Calculate $p(\sigma^2|\beta^1, \theta^1, \rho^0, z^0, y)$, we carry out a random draw to determine $\sigma^{2(1)}$.

$$\sigma^2|(\beta, \theta, \rho, z, y) \sim \mathcal{IG}(\alpha_0, \nu_0) \tag{15}$$

where $\alpha_0 = \frac{n}{2} + \alpha, \nu_0 = \nu + \frac{1}{2}\theta' S' S\theta$.

4. Calculate ρ using θ^1 and $\sigma^{2(1)}$ from previous steps. We represent the posterior distribution of ρ as follows,

$$p(\rho|\beta^1, \theta^1, \sigma^{2(1)}, z^0, y) \propto |I_n - \rho W| \cdot \exp\left\{-\frac{1}{2\sigma^2}\theta' S' S\theta\right\} \tag{16}$$

It is difficult to sampling above distribution. Therefore, we use Metropolis-Hastings algorithm with a random walk chain to generate draws (see Chib and Greenberg 1995). Let ρ^{old} denote the previous draw, and then the next draw ρ^{new} is given by:

$$\rho^{new} = \rho^{old} + c^*\phi \qquad \phi \sim \mathcal{N}(0, 1) \tag{17}$$

where c^* is called tuning parameter. The spatial term ρ is restricted $\lambda_{min}^{-1}, \lambda_{max}^{-1}$. Next, we evaluate the acceptance probability as follows,

$$\Psi(\rho^{old}, \rho^{new}) = \min\left(1, \frac{p(\rho^{new}|\beta^1, \theta^1, \sigma^{2(1)}, z^0, y)}{p(\rho^{old}|\beta^1, \theta^1, \sigma^{2(1)}, z^0, y)}\right) \tag{18}$$

Finally, we set $\rho = \rho^{new}$ with probability $\Psi(\rho^{old}, \rho^{new})$, otherwise $\rho = \rho^{old}$.

5. We sample z^1 draws from a truncated normal distribution using $\beta^1, \theta^1, \sigma^{2(1)}$ and ρ^1 as follows,

$$z_i|(\beta, \theta, \rho, \sigma^2, z_{-i}, y) \sim \begin{cases} \mathcal{TN}_{(0,\infty)}(x_i{}'\beta + \theta_i, 1) & if \quad y_i = 1 \\ \mathcal{TN}_{(-\infty,0]}(x_i{}'\beta + \theta_i, 1) & if \quad y_i = 0 \end{cases} \tag{19}$$

where $z_{-i} = (z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n)$.

We return to step 1 and repeat the sample with using $\beta^1, \theta^1, \rho^1, \sigma^{2(1)}$ and z^1 as the initial values. If the number of iterative times are enough large, the sampling values of each parameter become the draws from the true posterior distribution of the model.

3 Example of the Empirical Application

To illustrate the model in an applied setting we used data from a field survey in Indonesia which we conducted in 2008. Through systematic sampling, 500 households living at Toyomarto village (TY) and Candi Renggo village (CR), Singosari district are selected for the study. Data is collected employing the questionnaire interview survey method. The respondents are the husband, the wife or the head of family. The dependent variable was set to 1 for households which join HIPPAM (community based water supply system) and 0 for those which join PDAM (Indonesia local water company).

As explanatory variables, we used as following:

- *FAM*: number of people in household
- *GENDER*: dummy variable which equals 1 if respondent is male
- *AGE*: age of respondent
- *EDU*: dummy variable which is recorded as 1 if respondent has educational background in the level of elementary school or junior school and recorded as 0 if high school or university.
- *OCCU*: dummy variable which is coded as 1 if occupation of respondent is Agricuture or manufacturing and coded as 0 if service or unemployment
- *INCOME*: household's monthly income which is divided into 7 items (less than 0.5, 0.5-1.0, 1.0-1.5, 1.5-2.0, 2.0-2.5, 2.5-3.0, more than 3.0 million Rupiah), and we use the medians of each item.
- *LENGTH*: years of living in the area for respondent
- *COST*: water charge per day (Rupiah)

Table 1 shows the standard statistics of each variables.

Finally, in specifying the weight matrix, we reasoned that households with better community tie have ability to organize community based management system. Therefore, we define a spatial weight matrix using the data about community networks. Table 2 shows sample statistics of the community networks data. The data include information about what social groups does each household belong to, and we can calculate the social distance between household i and household j as follows:

$$\begin{aligned} w_{ij}^{gk} &= \begin{cases} 1 & \text{if household i and household j join the same social group k} \\ 0 & \text{otherwise} \end{cases} \\ w_{ij} &= \sum_{k=1}^{4} w_{ij}^{gk} \end{aligned} \tag{20}$$

The diagonal elements were all set to zero. Next we row standardize the matrix by dividing each element w_{ij} in the matrix by the row sum such that all rows sum to one. The row standardization does not change the relative social interaction among households. Other more complicated weighting schemes are possible, depending on how one wishes to quantify the degree of social interaction among households. For the purpose of this paper we simply want to account for social interaction effects in the decision to join HIPPAM, therefore any type of social interaction is acceptable.

In addition to the spatial probit model estimates, we also estimated a non-spatial probit model which does not include the spatial interaction term θ. Diffuse or conjugate priors were employed for all of the parameters β, σ^2 and ρ in both models. We iterate MCMC algorithm and sample 5000 parameters respectively and set 1000 samples as burn-in. The chain was considered to have practically converged after 1000 iterations based on a diagnostic proposed by Geweke (1992). The last 4000 draws were used to calculate the posterior mean and standard deviation of the parameters. Table 1 and 2 show the estimation results of each village.

Estimation results are summaries as follows. First, in both the villages, the results indicate very similar inferences would be drawn from the non-spatial probit model versus the spatial probit model. In addition, all of the estimated parameter ρ are negative and insignificant. Therefore, from this result, we cannot confirm the existence of the "social interaction effect" among households.

Second, the estimated parameter $LENGTH$ is positive and significant. This result indicates that the longer the respondents stay in the area of study the higher preference of them to join HIPPAM.

Third, we can find that the estimated parameter $COST$ is negative and significant. Then, even the price of monthly water usage is quite cheap, but this price in their point of view are one important demographic neighborhood to put into be consideration whether they have willingness to join or not.

4 Conclusion

In this paper, we show the spatial probit model with using Bayesian estimation method in order to investigate resident's spontaneous collaboration to manage community based water supply system. We describe the posterior distribution from the Bayes theorem and express the MCMC sampling method. Then, our approach applies to the empirical analysis of the data from a field survey in Indonesia.

From the estimation results, we can say that the length of living in the area and the price of monthly water usage have an important meaning for respondent to make a decision to join community based water supply system. Though, as for the social interaction, we do not have good result for parameter ρ yet. Therefore, we need use another approach to get weight matrix employing geographical neighbors data through Social Network Analysis in order to complete the previous analysis of demographic neighbors.

In this paper, we focused on the mechanism of resident's spontaneous collaboration to access water and did not take up the social and economic benefit of community based water supply system. Needless to say, however, it is important to clarify the flow of cost and benefit generated by community based water supply system. These are remained for the future works.

References

[1] Sun, D., R. K. Tsukawa and P. L. Speckman (1999), "Posterior Distribution of Hierarchical Models using CAR(1) Distribution," *Biometrica*, V86, 341-350.

[2] S. Chib and E. Greenburg. (1995), "Understanding the Metropolis-Hastings Algorithm", *American Statistical Association*, V46(4), 327-336.

[3] Smith, T. E., Lesage J. P. (2004), "A Bayesian Probit Model with Spatial Dependencies", *Advances in Econometrics*, V18.

[4] Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Means", *Bayesian Statistics 4*.

Table 1: Standard Statistics

Toyomarto(n=159)

Variable	Mean	Standard Deviation	Max	Min
FAM	3.925	1.082	7	2
GENDER	0.742	0.439	1	0
AGE	43.031	10.562	73	20
EDU	0.811	0.392	1	0
OCCU	0.547	0.499	1	0
INCOME	0.926	0.664	3.250	0.250
LENGTH	29.613	17.984	68	0.020
COST	264.921	642.547	7000	0

Candi Renggo(n=142)

Variable	Mean	Standard Deviation	Max	Min
FAM	3.711	1.275	7	1
GENDER	0.732	0.444	1	0
AGE	47.354	12.307	85	24
EDU	0.458	0.500	1	0
OCCU	0.225	0.419	1	0
INCOME	1.190	0.854	3.250	0.250
LENGTH	19.097	16.063	85	0.08000
COST	652.426	671.278	5000	0

Table 2: Community networks data

	1.Religious	2.Cultural/Social	3.Community organization	4.Finance
TY	138	28	14	15
CR	87	36	41	13

Table 3: Estimation Results(TY)

Toyomarto(n=159)

	Probit Model				
Variable	p.Mean	p.Std.Dev	90% Credible Interval		Geweke
constant	3.449	1.389	1.266	5.787	0.318
FAM	−0.734	0.178	−0.363	0.218	0.155
GENDER	−0.058	0.468	−0.838	0.697	0.408
AGE	−0.029	0.021	−0.063	0.005	1.097
EDU	−0.069	0.533	−0.982	0.794	0.402
OCCU	−0.289	0.428	−1.007	0.383	0.236
INCOME	0.252	0.363	−0.326	0.876	1.075
LENGTH	0.024	0.013	0.002	0.046	0.476
COST	−0.004	0.001	−0.005	−0.003	0.390

	Spatial Probit Model				
Variable	p.Mean	p.Std.Dev	90% Credible Interval		Geweke
constant	3.620	1.545	1.212	6.200	1.872
FAM	−0.085	0.204	−0.415	0.246	0.603
GENDER	0.035	0.509	−0.801	0.861	0.607
AGE	−0.031	0.023	−0.069	0.006	1.680
EDU	−0.136	0.579	−1.112	0.788	0.322
OCCU	−0.234	0.485	−1.062	0.561	2.863
INCOME	0.228	0.367	−0.364	0.844	0.499
LENGTH	0.026	0.014	0.003	0.049	1.949
COST	−0.004	0.001	−0.005	−0.003	2.224
σ^2	0.128	0.027	0.092	0.177	1.220
ρ	−9.016	5.486	−17.345	0.175	1.203

Table 4: Estimation Results(CR)

Candi Renggo(n=142)

	Probit Model				
Variable	p.Mean	p.Std.Dev	90% Credible Interval		Geweke
constant	−1.272	0.749	−2.501	−0.045	1.634
FAM	0.199	0.119	0.006	0.394	1.360
GENDER	0.554	0.351	0.005	1.144	0.003
AGE	−0.026	0.014	−0.048	−0.004	0.787
EDU	0.865	0.340	0.312	1.446	0.740
OCCU	1.059	0.363	0.478	1.661	1.053
INCOME	0.129	0.181	−0.168	0.426	0.059
LENGTH	0.039	0.011	0.020	0.058	0.831
COST	−0.001	0.0002	−0.001	−0.0003	1.147

	Spatial Probit Model				
Variable	p.Mean	p.Std.Dev	90% Credible Interval		Geweke
constant	−1.376	0.800	−2.714	−0.092	0.509
FAM	0.196	0.123	0.002	0.404	0.711
GENDER	0.594	0.370	0.001	1.216	3.406
AGE	−0.026	0.014	−0.050	−0.004	0.519
EDU	0.939	0.362	0.351	1.544	0.578
OCCU	1.157	0.397	0.497	1.819	1.683
INCOME	0.149	0.193	−0.170	0.476	0.544
LENGTH	0.041	0.013	0.021	0.063	0.866
COST	−0.001	0.0002	−0.001	−0.0004	0.903
σ^2	0.127	0.026	0.091	0.175	0.274
ρ	−5.092	3.222	−10.121	0.067	0.412