Spatial modelling of households' knowledge about arsenic pollution in South Asia: Evidence from Bangladesh

Abstract

Households' knowledge about arsenic threat from drinking water is an important issue for public health and policy implication of Bangladesh. In this study, I use spatial statistical models to investigate the determinants and spatial dependence of households' knowledge about arsenic risk. The binary joins matrix/ binary contiguity matrix and the inverse distance spatial weights matrix techniques are used to capture spatial dependence in the data. My analysis extends the spatial model by allowing spatial dependence to vary across the divisions and regions. I find positive spatial correlation in households' knowledge across neighboring districts at district, divisional and regional levels but the strength of this spatial correlation varies considerably based on spatial weight. I also find that the literacy rate, daily wage rate of agricultural labour, arsenic status, percentage of red mark tube well of a district contribute positively significantly to the households' knowledge. These findings have policy implications both at regional and national level for taking appropriate steps in mitigating the present arsenic crisis and to ensure arsenic free water in Bangladesh.

Keywords: Spatial modeling; Households' knowledge; Arsenic pollution; Bangladesh

1. Introduction

Econometric models taking into account spatial interactions among economic units have been increasingly used by economists over the last several years and some important advances have been done in both theoretical and empirical studies. Spatial econometrics is becoming more popular in many scientific fields including social sciences. More recently, spatial interaction has increasingly received more attention in mainstream econometrics as well, both from a theoretical as well as from an applied perspective (Anselin, 2001; Anselin, 2002; Anselin & Bera, 1998). Spatial econometric techniques have been developed to effectively capture spatial processes in natural or experimental data (Anselin, 1988, 2001; Haining, 1990; Coughlin & Garrett, 2004). Spatial dependence not only means lack of independence between observations, but also a spatial structure underlying these spatial correlations (Anselin & Florax, 1995).

The modeling of spatial variation of arsenic threat in south Asia is a very important issue for a number of reasons. Natural arsenic (As) pollution of drinking water supplies has been reported from over 70 countries, a serious health hazard to an estimated 150 million people world-wide (Brammer & Ravenscroft, 2009). The high arsenic groundwater in Asia has become a priority health issue. Around 110 million of those people live in ten countries in South and South-East Asia: Bangladesh, Cambodia, China, India, Laos, Myanmar, Nepal, Pakistan, Taiwan and Vietnam. In South Asia, Bangladesh is the most arsenic affected country and it has been recognized as a big threat and challenge to public health. World Bank, UNICEF, WHO agreed that Bangladesh is in dire straits regarding the arsenic problem (Sharma, 2009). It is estimated that more than 65% of the population of Bangladesh live in arsenic contaminated areas (Chowdhury, 2001) and are at risk of drinking water containing >50μg/L (Ng & Moore, 2005). Approximately 20 million people have already developed signs of arsenicosis. As a therapeutic measure, selenium has been reported to counteract arsenic toxicity (Wang, 2001). A good numbers of research are being carried out on the cause, characteristics and consequences of arsenic pollution on skin lesion. Only a few numbers of studies have been done on social and economic point of view (Caldwell, Mitra & Smith, 2003; Hanchett, 2004; Paul, 2004; Hassaan, Atkins & Dunn, 2005; Nahar, Hossain & Hossain, 2007; Sarker, 2008) however, it is important to note that these studies ignore any spatial patterns that may be present in the data although it is a very crucial and critically important issue. When the spatial dependence is ignored, the OLS estimates will be inefficient, the t and F-statistics for tests of significance will be biased and the R^2 measure of fit will be misleading (LeSage & Pace, 2004, Anselin, 1988). In other words, the statistical interpretation of the regression model will be wrong. However, the OLS estimates themselves remain unbiased, contrary to what is sometimes suggested in the literature. Since the attributes associated with the built environment and natural amenities are spatially located, it is reasonable to hypothesize that health disorders like arsenicosis are spatially clustered based on neighboring socioeconomic, demographic and environmental attributes. While the application of explicit spatial econometric methods has recently shown a tremendous increase in the social sciences in general and economics in particular (Anselin, 2001), to date, there have been only a few of studies that employed spatial regression analysis in the study of health related data in South Asia. It is very important for us to identify the efficient and consistent influencing factors of the arsenic related issues in the arsenic affected areas at first, before planning of any mitigation methods and intervention programs for arsenic free water. With this background, my study is focused on detecting the spatial dependence and investigating the determinants of households' knowledge about arsenic contaminated drinking water. This paper is important for several reasons. First, the present study is the first to address the spatial dimension for environmental health problem awareness in Bangladesh, second, this methodology provides consistent and efficient estimation and third, my results have important implications for policy both at the regional and national level, especially those involving the design of regional coordination for arsenic free drinking water.

2. Methods

2.1 Data Sources

The data used in this paper was collected from the six sources: (i) Bangladesh Multiple Indicator Cluster Survey (MICS) 2006 (BBS & UNICEF, 2007) for percentage of households who have known of arsenic contamination drinking water and percentage of red marks tube well: MICS 2006 was conducted in 1,950 primary sampling units and covered as many as 62,463 households throughout the country during June through October, 2006, (ii) Statistical Pocketbook of Bangladesh (BBS, 2008) for average household size of different districts, (iii) Statistical Yearbook of Bangladesh (BBS, 2008) for literacy rate, (iv) Population Census 2001 for population density characteristics of different districts, (v) Directorate of Agricultural Extension personnel for daily average wage rate of agricultural labour and (vi) finally, aggregate data set of Groundwater Studies for Arsenic Contamination in Bangladesh for determining the arsenic risk status of different districts which was conducted by British Geological Survey (BGS) and the Department of Public Health Engineering (DPHE) (BGS/DPHE, 2001), Bangladesh. The final data set for this survey consisted of samples from 3,534 tubewells from 61 of 64 districts and from 433 of the 496 upazilas of Bangladesh. The sampled area was approximately 129,000 km^2, compared with a total area for Bangladesh of about 152,000 km^2.

2.2 District model

A spatial lag model is a formal representation of the outcome of processes of social and spatial interaction. Spatial lag or regressive spatial autoregressive model includes a spatial lagged dependent variable on the right hand side of the regression specification (Anselin, 1988). A second approach to spatial autoregressive process or spatial autoregressive modeling is known as the spatial error model. Examining the spatial effects on households knowledge about arsenic threat in drinking water of Bangladesh, I used the basic model of spatial correlation developed by Anselin (1988), Cliff & Ord (1981) and Coughlin & Garrett (2004) allows for spatial dependence in the dependent variable or in the error component.

$$y = \rho Wy + X\beta + \varepsilon \qquad \text{.......(1)}$$

$$\varepsilon = \lambda W\varepsilon + u \qquad \text{.......(2)}$$

Where X is an (n x (k+1)) matrix of observations on the explanatory variables, y and ε are n x 1 vectors and β is a (k+1) vector. If E (εε) = $\sigma^2 I$, where I is the nxn identity matrix then the arrangement properties of the attributes are irrelevant to the specification of the model. Wy and $W\varepsilon$ are exogenously specified nxn weights matrices, lag and error, respectively. The scalar ρ and λ are spatial lag and spatial error coefficients, respectively.

The following district model was used for detecting the spatial effect:

$$y = \rho Wy + \beta_0 + \beta_1 hhsize + \beta_2 lrate + \beta_3 pdensity + \beta_4 dwrate + \beta_5 asstatus + \beta_6 rmtwell + \varepsilon, where, \varepsilon = \lambda W\varepsilon + \mu \quad \ldots (3)$$

where y is the percentage of households who have known about arsenic contaminated drinking water for a district of Bangladesh, *hhsize* is the house hold size, *lrate* is literacy rate, *pdensity* is the population density in per square kilometer, *dwrate* is daily wage rate of agricultural labour and *asstatus* is the binary variable taking value 1 for the districts that are under arsenic risk (average arsenic concentration >0.10µg/L) according DPHE/BGS survey and 0 otherwise, *rmtwell* is percentage of red marks tube well in a district.

2.2.1 Spatial weights matrix (W)

The spatial weights matrix W is a N X N positive squared, non-stochastic matrix, whose elements w_{ij} show the intensity of interdependence between the spatial units i (in the row of the matrix) and j (column) in which the rows and columns correspond to the cross-sectional observations. A weight matrix summarizes the spatial relationships in the data. Many alternative weighting schemes for W have been used in the literature. It is important to note that the elements of the weights matrix are exogenous to the model; that is, it is assumed that the researcher knows how the observations are related to each other. Typically, they are based on the geographic arrangement of the observations or contiguity. The cross-sectional spatial weights matrix formalizes the potential correlation among districts for which many alternative weighting representations are possible. I consider the binary join matrix and the inverse distance spatial weights matrix specifications of W in my empirical models to highlight any differences in spatial patterns of household knowledge.

2.2.1.1 Binary contiguity matrix

One of the most common weights matrix in spatial econometrics literature is the binary joins matrix/ binary contiguity matrix (Anselin, 1988; Case, 1992; Cliff & Ord, 1981; Coughlin & Garrett, 2004) where $w_{ij} = 1$ if observations i and j ($i \neq j$) share a common border, and $w_{ij} = 0$ otherwise. Consequently, two spatial units are either neighbors or are not, hence the use of the term binary contiguity. By convention, the diagonal elements of *W* are set to zero (implying that locations are not neighbors of themselves). Because the weight matrix shows the relationships between all of the observations, its dimension is always N X N, where N is the number of observations.

$$W = \begin{bmatrix} 0, w_{12} \cdots\cdots\cdots\cdots w_{1j} \\ w_{21}, 0 \cdots\cdots\cdots\cdots w_{2j} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ w_{i1}, w_{i2} \cdots\cdots\cdots\cdots 0 \end{bmatrix} \quad \ldots\ldots(4)$$

A limitation of the binary joins matrix is that it assumes equal weights across all bordering spatial neighbors and does not allow the effective capture of spatial distances across all cross sectional. For the estimation of spatial regression models, I used row-standardized binary spatial weights matrix for meaningful interpretation of the results. Row normalizing gives the weight matrix some nice theoretical properties. The row standardization consists of dividing each element in a row by the corresponding row sum, hence it is effectively including a weighted average of neighboring values into the regression equation. In this specification, the elements of matrix W are row-standardized by dividing each w_{ij} by the sum of each row i. The resulting row-standardized weights matrix is likely to become asymmetric, even though the original matrix may have been symmetric.

$$W = \begin{bmatrix} 0, w_{12} / \sum w_{1j} \cdots\cdots\cdots w_{1j} / \sum w_{1j} \\ w_{21} / \sum w_{2j}, 0 \cdots\cdots\cdots w_{2j} / \sum w_{2j} \\ \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \\ \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \\ \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \\ w_{i1} / \sum w_{ij}, w_{i2} / \sum w_{2j} \cdots\cdots\cdots 0 \end{bmatrix} \cdots\cdots\cdots\cdots (5)$$

2.2.1.2 Inverse distance matrix

A limitation of the binary joins matrix is that it assumes equal weights across all bordering spatial neighbors and does not allow the effective capture of spatial distances across all cross sectional units. Thus, I also considered the various measures of spatial distance that have been discussed in the literature (Bodson & Peeters, 1975; Coughlin & Garrett, 2004; Dubin, 1988; Garrett & Marsh, 2002; Hern´andez-Murillo, 2003). Measures of spatial contiguity include the distance specification between districts, where $w_{ij} = 1/d_{ij}$, the inverse distance squared, where $w_{ij} = 1/ d^2_{ij}$ and exponential distance decay, where $w_{ij} = \exp(-d_{ij})$. As the distance between districts i and j increases (decreases), w_{ij} decreases (increases), thus giving less (more) spatial weight to the district pair when $i \neq j$. In all cases, $w_{ij} = 0$ for $i = j$. While there is no consensus on how distance between cross-sectional units should be measured, I follow Hern´andez-Murillo (2003) and Coughlin & Garrett (2004) and consider the distance between district population centers.

2.3 Relation between spatial lag (ρ) and spatial error (λ) coefficients

The scalar ρ is the spatial lag coefficient, positive spatial correlation exists if $\rho > 0$, negative spatial correlation if $\rho < 0$, and no spatial correlation if $\rho = 0$ and λ is the spatial error coefficient. The errors are positively correlated if $\lambda > 0$, negatively correlated if $\lambda < 0$, and spatially uncorrelated correlated if $\lambda = 0$. The cases under the consideration of the models (1) and (2) are following:

Case (i): ρ = 0, λ = 0

Under this assumption, spatial lag weight matrix component pWy and spatial error weight matrix component $\lambda W\varepsilon$ will be zero and estimating the parameter β, then the general spatial model (1) and (2) reduce to the standard regression model. There is no spatial lag or serial error correlation in this model.

Case (ii): ρ≠ 0, λ = 0

The second case focuses on the spatial lag and ignoring the presence of spatial error correlation. If setting $\lambda = 0$ and estimating (ρ, β), it is spatial lag model which is clearly more meaningful. The general spatial model equation (1) and (2) tends to spatial lag model (1). In such models, the dependent variable in location i is not only determined by covariates (X) specific to location i, but also by the value of the same dependent variable at other locations.

Case (iii): ρ = 0, λ≠ 0

The third case focuses on the error term and ignoring the presence of spatial lag correlation. If ρ= 0 and estimating the other parameters λ and β, then the general spatial model (1) and (2) take the equation (2) form.

Case (iv): ρ ≠ 0, λ≠ 0

The fourth case considers the both spatial lag and spatial error term. In spatial lag + error model all three parameters (ρ, λ and β) are estimated. These models may be viewed in practice as resulting from poorly specified lag matrices Wy which results in spatial interactions in the error terms that need to be taken into account Wε (Anselin & Bera, 1998). The most general spatial lag model and error model under the case four takes the following form;

$$y = pWy + X\beta + \lambda W\varepsilon + \mu \quad \text{......(6)}$$

2.4 Divisional/Regional Model

The basic spatial model detailed above assumes that the influence of spatial dependence is the same for all districts. To reveal differences in spatial correlation for geographic regions, I modify equations (1) and (2) to allow for different spatial correlation coefficients in four divisions and two regions of Bangladesh. I use four ex-administrative divisions in the contiguous 64 districts. The spatial model with regional/divisional spatial correlation coefficients may be written as:

$$y = \sum_{k=1}^{D} \rho_k W_k y + X\beta + \varepsilon \quad \text{...... (7)}$$

where, $$\varepsilon = \sum_{k=1}^{D} \lambda_k W_k + u = \left(\mathrm{I} - \sum_{k=1}^{D} \lambda_k W_k \right)^{-1} u \quad \text{......(8)}$$

Here, D denotes the total number of divisions/regions and ρ_k and λ_k denote the spatial lag and spatial error lag coefficients, respectively, for division k. Each coefficient, ρ_k, measures the average correlation between a district in region k and the spatially weighted household knowledge of all other districts. W_k remains the (N × N) spatial weights matrices w_k. Each matrix W_k is constructed by pre-multiplying by a dummy variable that equals unity if district *i* is located in division/region *k*, and zero otherwise. In the case of a contiguity matrix, household knowledge in district *i* located in region *k* to be affected by household knowledge of all districts *j* that border district *i*, regardless of whether district *j* is in the same division/region as district *i*.

2.5 Model comparison

To compare models, I consider the Akaike`s Information Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Schwartz, 1978) based on the ML method. AIC, a penalized log likelihood criterion is defined by

$AIC = -2\,\ell + 2K$……………………………………………………………..(9)

Where ℓ is the log likelihood and k is the number of parameters.

$SC = -2\,\ell + K \ln (n)$ ……………………………………………………………….. (10)

Where n is the number of observations.

In theory, the lower AIC and SC/BIC are the better specification.

3. Results and Discussion

Table 1 provides descriptive statistics for the variables employed in the models. These statistics provide mean, maximum and minimum values and standard deviation. The spatial relationship among locations in a spatial framework is often modeled via a contiguity matrix.

3.1 Spatial estimates with binary join weights

Table 2 presents the spatial estimation results. Column [1] in table 2 corresponds to a standard regression model where no spatial effects among the neighboring districts are taken into account. The coefficients represent the effects of the explanatory variables on household knowledge about arsenic pollution. Column [2] corresponds to a specification where the spatial interaction among the districts is accounted. This model suggests that the knowledge of households about arsenic risk depends positively on neighboring districts households' knowledge. Columns [2], [3] and [4] present the results from three alternative specifications in which I account for spatial effect using binary join spatial weights to determine contiguity among the districts. As I discussed before, this scheme considers as neighbors only those districts that are adjacent to each other. Column [2] corresponds to the model with a spatial lag in the dependent variable, column [3] corresponds to the model with a spatial dependence in the error term and column [4] corresponds to the model with a spatial lag and error term. As I can see from the table, the spatial dependence coefficients are statistically significant. In particular, the spatial lag coefficient ρ indicates the presence of interaction among the districts. Specifications [1] through [4] suggest that the literacy rate of districts has a positive impact on the knowledge of arsenic pollution of households. The same is true for daily wage rate of agricultural labour and arsenic status of a district. Household size and population density per square kilometer have negative impact on the knowledge of arsenic pollution of households but the household size is statistically significant in all models and population density per square kilometer is not significant in case of spatial lag and spatial lag & error model. In the models corresponding to the binary spatial weights, the spatial lag in the dependent variable is statistically significant from zero at the 1 percent level and the spatial lag in the error term is not statistically significant. The results from a join spatial model (spatial lag +error) in column [4]. The numerical interpretation of the estimated coefficients is as follows. The findings in column [2] of table 2, for example, suggest that a 1% increase the literacy rate, on average, an increase of about 0.20 percentage households' knowledge about arsenic pollution. A decrease of household size, 1 person per household induces increase 5.43 percentage households' knowledge about arsenic pollution. The interpretation for the dummy variable is straightforward, as the coefficient indicates that the district arsenically not safe the households who know about arsenic pollution is about 7.14 percent higher, on average, than the districts which is arsenically safe. The results from the models that both the spatial lag (column 2) and error (column 3) models reveal that the spatial lag and error coefficients are significantly different from zero. However, the Akaike Information Criterion (AIC), Schwarz Criterion (SC) and the log-likelihood statistics all reveal that the spatial lag model presented in columns 2 is provided a better fit than the spatial error model. The results from the model that include both the spatial lag and error term (column 4) reveals that only the spatial lag coefficient is significantly different from zero. The findings are presented in table 2 suggest that spatial dependence in district households' knowledge may be the best model using a spatial lag. This is supported by the AIC and SC, which are directly comparable across models and weigh the explanatory power of a model (based on the maximized value of the log-likelihood function).

3.2 Spatial estimates with inverse distance spatial weights

An alternative definition of neighborhood effects in the households' knowledge about arsenic pollution allows knowledge of household in nearby districts that are not necessarily adjacent to affect a specific district. In this case, the use of inverse-distance spatial weights is more appropriate to identify spatial interactions among the distance spatial weights of districts. Columns [2] through [4] in table 3 present the estimation results using spatial weights computed as the inverse distance between districts' population centers. The results are qualitatively similar to those in table 2. The results from a joint spatial model, in column [4], suggest spatial interaction in both the dependent variable and in the error term. The coefficients for the spatial lag in columns [2] and [4] are positive, supporting the conclusion that there is a positive interaction in the households' knowledge among the districts. The spatial coefficients are statistically significant in model 2 and model 4 but the spatial error coefficients are statistically insignificant in model 3 and model 4. The AIC, BIC and log likelihood statistics suggest that the spatial models show better fit than non-spatial model (column1of table 3). Among the four specifications, log likelihood value is lowest for join (spatial lag + error) model (column 4) and AIC and SC are lowest in spatial lag model (column 2). The findings from table 3 suggest that the spatial lag model is the best model for expressing the spatial dependence of district households' knowledge.

3.3 Divisional and regional spatial estimates

Table 4 presents the results from four specifications that allow for regional and divisional differences in the spatial correlation coefficients by using binary joins and inverse distance contiguity. Because my results in table 2 and table 3 indicate that a spatial lag models are the appropriate specifications, therefore, I only consider for divisional and regional differences in spatial lag coefficients among the four specifications. Column 1 reports the estimates from the spatial lag model that allows for regional-specific spatial lag coefficients at the administrative division level, while Column 3 allows regional-specific spatial lag coefficients at the regional (North, South) level when neighbors are defined as common border. I find considerable evidence that the effects of spatial correlation in the dependent variable vary by division. As reported in column [2] of table 4 two regional correlation coefficients are positive and statistically significant at the 1 percent level; furthermore, visual inspection of the estimated coefficients suggests differences in the magnitude of spatial correlation between a district in a given region and all other districts. The regression that allows for division level spatial coefficients (column 1 of table 4) provides a same picture with the region level model (column 3). Estimates of positive and significant spatial correlation in four divisions range from 0.6923 to 0.6236 when neighbors are defined by binary joins matrix and 0.3760 to 0.3166 when spatial weight are defined by inverse distance matrix. Estimates of spatial correlation in two regions are 0.6197 and 0.6526 when neighbors are defined by binary joins matrix and 0.3332 and 0.3328 when spatial weight are defined by inverse distance matrix. All spatial coefficients for regional are positive and statistically significant ($P<0.01$). The AIC, BIC and log likelihood statistics suggest that the binary joins matrix contiguity spatial models show better fit than inverse distance spatial weights matrix models in both regional and divisional models.

3.4 Equality of spatial coefficients of divisional and regional models

The spatial coefficients in the Dhaka, Chittagong, Khulna and Rajshahi divisions as well as Northern and Southern regions of Bangladesh are substantially larger when neighbors are defined by common border as opposed to inverse distance.

The results from the divisional and regional specific models in table 4 suggest that the spatial effects on district households' knowledge are more or less homogeneous across divisions and regions. Further evidence is reported in table 5 on basis of table 4, where I present results the seven pair wise hypothesis tests of the equality of the spatial correlation coefficients when neighbors defined as common border from the fourteen possible pair wise equality tests at the divisional level and regional level. There are six pair wise equality tests for model 1 at the divisional level and one pair wise equality test for model 3 at the regional level of table 4. Using the common-border (binary join matrix) neighbor definition, the test results are shown that the spatial correlation for districts in each division is statistically homogenous from the correlation in other divisions, with the exception of the Rajshahi and Khulna divisions. The test results (column 2 and 4 of table 4) indicate that spatial correlations are not significantly different across all divisions and regions when neighbors are defined by the inverse distance.

4. Conclusions

In this study, I estimate spatial econometric models to explore the spatial dependence in the knowledge of households about arsenic pollution like environmental health problem. This study is the first to directly model and provide estimates on the spatial interdependence of a district households' knowledge about arsenic contaminated water and the approach taken here affords the estimation of consistent and efficient coefficients. The results from spatial models strongly indicate that the households' knowledge about arsenic contaminated drinking water are combination of the household characteristics, arsenic related factors of individual districts and the households' knowledge of their neighbors districts. Five characteristics- literacy rate, daily wage rate of agricultural labor, household size, arsenic status and percentage of red mark tube well are significantly related to household knowledge of a district. Based on the AIC, SC and log likelihood, all of the spatial models are preferred to the without spatial weight specification, but the spatial lag model that utilizes the binary joins contiguity weights matrix provides the best fit of the data. The models in which I assume a common spatial lag coefficient for all districts, this results indicate that one percentage increase in the average households' knowledge of neighboring districts generates between a 0.33 and 0.61 percentage increase of a given district households' knowledge, depending on the specification. Using either a binary or inverse distance weights matrix in the estimation of spatial effects, this results provide strong evidence that significant spatial correlation exists in district, divisional and regional level models. These results suggest that district should pay particular attention to policies in neighboring districts and policy maker should realize that improving the households' knowledge level in neighboring districts are likely to affect households' knowledge in their own district, therefore, a key issue for policy development is how to stimulate educational attainment, promote daily wage of agricultural labor and decrease household size and population density could increase household knowledge and result in sustainable development and poverty alleviation of regions that are both knowledge on arsenic pollution and economically lagging. This needs to be addressed both in terms of national level policies and more emphatically within regional and sub-regional development strategies than it has been hither.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Aut Cont*, 19, 716–723.

Anselin, L. (1988). *Spatial Econometrics Methods and Models*. New York: Kluwer Academic.

Anselin, L. (1995). *SpaceStat. A Software Program for the Analysis of Spatial Data,* Version 1.80. Regional Research Institute, West Virginia University, Morgantown, West Virginia.

Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, 97, 113-139.

Anselin, L. (2002). Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models. *Journal of Agricultural Economics*, 27, 247-267.

Anselin, L., & Bera, A. (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, In A. Ullah & D. Giles (Eds.) *Handbook of Applied Economic Statistics*, New York.

Anselin, L., & Florax, R. (1995). Introduction. In L. Anselin & R. Florax (Eds.) *New Directions in Spatial Econometrics* (pp. 3–18). Berlin: Springer-Verlag.

BBS. (2003). *Population Census 200, National Report (Provisional)*, Ministry of Planning, Government of the People's Republic of Bangladesh, Dhaka: Bangladesh Bureau of Statistics.

BBS. (2008). *Statistical Pocket Book of Bangladesh*. Ministry of Planning, Government of the People's Republic of Bangladesh. Dhaka: Bangladesh Bureau of Statistics.

BBS. (2008). *Statistical Yearbook of Bangladesh*. Ministry of Planning, Government of the People's Republic of Bangladesh. Dhaka: Bangladesh Bureau of Statistics.

BBS-UNICEF. (2007). *Multiple Indicators Clusters Survey Bangladesh 2006 (Key* *findings)*, BBS-UNICEF, Dhaka: Bangladesh Bureau of Statistics.

BGS and DPHE. (2001). *Arsenic Contamination of Groundwater in Bangladesh. Volume 2: final report.* In D. G. Kinniburgh, & P. L. Smedley (Eds.), BGS technical report WC/00/19. Key worth: British Geological Survey and Department of Public Health Engineering, Ministry of Local Government, Rural Development and Co-operatives, Government of Bangladesh. (Retrieved February 27, 2007 from http://www.bgs.ac.uk/arsenic/bangladesh/reports.htm)

Bodson, P., & Peeters, D. (1975). Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to a Belgian Labor Demand Function. *Environment and Planning A*, 7, 455-472.

Brammer, H., & Ravenscroft, P. (2009). Arsenic in groundwater: A threat to sustainable agriculture in South and South-east Asia. *Environment International*, 35, 647-654.

Caldwell, B. K., Caldwell, J. C., Mitra, S. N., & Smith, W. (2003). Searching for an optimum solution to the Bangladesh arsenic crisis. *Social Science & Medicine*, 56, 2089–2096.

Case, A. (1992). Neighborhood Influence and Technological Change. *Regional Science and Urban Economics,* 22, 491-508.

Chowdhury, S. (2001). Afflicted with arsenic. *The Daily Star Weekend Magazine (Dhaka)*, January 26.

Cliff, A., & Ord, J. (1981). *Spatial Processes: Models and Application*. London: Pion Ltd..

Coughlin, C., & Garrett T. A. (2004). *Spatial Probit and the Geographic Patterns of State Lotteries.* Working Paper, 2003-042B, Federal Reserve Bank of St. Louis.

Dubin, R. A. (1988). Estimation of Regression Coefficients in the Presence of Spatially

Autocorrelated Error Terms. *Review of economics and statistics*, 70, 466-474.

Garrett, T. A., & Marsh, T. L. (2002). The Revenue Impacts of Cross-Border Lottery Shopping in the Presence of Spatial Autocorrelation. *Regional Science and Urban Economics,* 32, 501-519.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.

Hanchett, S. (2004). *Social aspects of the arsenic contamination of drinking water: A review of knowledge and practice in Bangladesh and West Bengal.* Government of the People's Republic of Bangladesh.: Report Prepared for the Arsenic Policy Support Unit, Ministry of Local Government, Rural Development & Cooperatives.

Hassan, M. M., Atkins, P., & Dunn, C. (2005). Social implications of arsenic poisoning in Bangladesh. *Social Science & Medicine*, 61, 2201–2211.

Hern´andez-Murillo, R. (2003). Strategic Interaction in Tax Policies Among States. *Federal Reserve Bank of St. Louis Review*, 85(3), 47-56.

Lesage, J. P., & Pace, R. K. (2004). Introduction to Advances in Econometrics. In J. P. Lesage, & R. K. Pace (Eds.), *Spatial and Spatiotemporal Econometrics* 18 (pp1-32). Oxford: Elsevier Ltd.

Nahar, N., Hossain, F., & Hossain, M. D. (2008). Health and Socio-Economic Effects of Groundwater Arsenic Contamination in Rural Bangladesh: New Evidence from Field Surveys. *Journal of Environmental Health*, 70(9), 42-47.

Ng, J. C., & Moore, M. R. (2005). Arsenic in drinking water: a natural killer in Bangladesh and beyond. *Medical Journal of Australia*, 183, 562-563.

Paul, B. K. (2004). Arsenic contamination awareness among the rural residents in Bangladesh. *Social Science & Medicine*, 59, 1741-1755.

Sarker, M. M. R. (2008). Determinates of Arsenicosis Patients Treatment Cost in Rural Bangladesh. Bangladesh. *Journal of Environmental Science*, 14(1), 80-83.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

Wang, W., Wang, L., Shaofan, H., & Tan, J. (2001). Prevention of endemic arsenism with selenium. *Current Science*, 81, 215-1218.

Table 1. Descriptive Statistics of Variables

Variables	Mean	Maximum	Minimum	Standard Deviation
Households who have knowledge about arsenic contamination (%)	77.98	98.50	30.80	18.48
Household size	4.74	5.94	3.85	0.46
Literacy rate (%)	43.78	65.90	28.00	8.52
Population density (per sq. km)	951.64	5857.60	65.39	732.11
Daily wage rate of agricultural labour (TK.)	185.31	250.00	130.00	12.22
Arsenic status	0.72	1	0	0.45
Red mark tube well (%)	10.03	10.59	0	49.52

Table 2 Spatial estimates with binary weights of the knowledge of households about arsenic contaminated drinking water

Variable	Coefficients	
	No Spatial Effect (1)	Spatial Lag (2)
Constant	15.1360 (14.5759)	8.7413 (11.2115)
Literacy rate (%)	0.3059** (0.1418)	0.1986** (0.1004)
Daily wage rate of agricultural labour (TK.)	0.3918*** (0.0567)	0.1770*** (0.0540)
Population density (per sq. km)	-0.0040** (0.0016)	-0.0015 (0.0013)
Household size	-6.6984*** (2.3774)	-5.4292*** (1.8318)
Arsenic status (Dummy)	12.27132*** (3.0532)	7.1476*** (2.4615)
Red mark tube well (%)	0.3576*** (0.1090)	0.2079** (0.0864)
Rho	--	0.6108*** (0.0911)
Log Likelihood	-224.6958	-207.6679
AIC	465.3916	433.3358
BIC	482.6626	452.7657

Standard error in the parentheses. Asterisks ** and *** indicate statistical significance at the 5% and 1% levels, respectively.

Table 3: Spatial estimates with inverse-distance weights of the knowledge of households about arsenic contaminated drinking water

Variable	Coefficients	
	No Spatial Effect (1)	Spatial Lag (2)
Constant	15.1360 (14.5759)	8.1300 (13.4457)
Literacy rate (%)	0.3059** (0.1418)	0.2718** (0.1298)
Daily wage rate of agricultural labour (TK.)	0.3918*** (0.0567)	0.3015*** (0.0576)
Population density (per sq. km)	-0.0040** (0.0016)	-0.0042*** (0.0015)
Household size	-6.6984*** (2.3774)	-4.9151** (2.2261)
Arsenic status (Dummy)	12.27132*** (3.0532)	9.3102*** (2.9065)
Red mark tube well (%)	0.3576*** (0.1090)	0.3191*** (0.1000)
Rho	--	0.3332*** (0.0930)
Log Likelihood	-224.6958	-218.8461
AIC	465.3916	455.6921
BIC	482.6626	475.1221

Standard error in the parentheses. Asterisks ** and *** indicate statistical significance at the 5% and 1% levels, respectively.

Table 4: Divisional and regional spatial estimates of the knowledge of households about arsenic pollution

Variables	Coefficients	
	Divisional Spatial Lag	Regional Spatial Lag

	Binary join (1)	**Inverse distance** (2)	**Binary join** (3)	**Inverse distance** (4)
Constant	-1.8912	2.5762	1.0168	8.2198
	(14.2902)	(17.5653)	(12.5161)	(15.6276)
Literacy rate (%)	0.2549**	0.2196**	0.2448**	0.2792**
	(0.1210)	(0.1078)	(0.1161)	(0.1403)
Daily wage rate of agricultural labour (TK.)	0.1826***	0.3260***	0.1812***	0.3014***
	(0.0574)	(0.0631)	(0.0534)	(0.0581)
Population density (per sq. km)	-0.0023	-0.0036**	-0.0016	-0.0042***
	(0.0014)	(0.0017)	(0.0013)	(0.0015)
Household size	-4.6135**	-4.2527	-4.8057**	-4.9236**
	(2.3269)	(2.778)	(1.8682)	(2.3505)
Arsenic status (Dummy)	8.1779***	8.7845***	7.0255***	9.3110***
	(2.4742)	(2.9862)	(2.4305)	(2.9075)
Red mark tube well (%)	0.1666**	0.3287***	0.2168**	0.3190***
	(0.0819)	(0.1149)	(0.0856)	(0.1007)
ρ_1 (Dhaka)	0.6532***	0.3166***		
	(0.1001)	(0.1030)		
ρ_2 (Chittagong)	0.6547***	0.3448**		
	(0.1108)	(0.1431)		
ρ_3 (Khulna)	0.6236***	0.3685***		
	(0.0932)	(0.1097)		
ρ_4 (Rajshahi)	(0.6923)***	0.3760***		
	(0.1025)	(0.1106)		
ρ_1 (North)			0.6526***	0.3328***
			(0.0953)	(0.0995)
ρ_2 (South)			0.6197***	0.3332***
			(0.0901)	(0.0931)
Log Likelihood	-206.1423	-217.9901	-206.8093	-218.846
AIC	436.2846	459.9802	433.6186	457.692
BIC	462.1912	485.8868	455.2074	479.2808

Standard error in the parentheses. Asterisks ** and *** indicate statistical significance at the 5% and 1% levels, respectively.

Table 5: Spatial coefficients equality with binary join weights for divisions and regions

	Dhaka	Chittagong	Khulna	Rajshahi	North	South
Dhaka	----	ns	ns	ns		
Chittagong	ns	----	ns	ns		
Khulna	ns	ns	----	*		
Rajshahi	ns	ns	*	----		
North					----	ns
South					ns	----

Asterisk * indicates statistical significance at the 10% and ns means statistical insignificance.