Spatial Effects of Nutrient Pollution on Drinking Water Production[1]

Roberto Mosheim
Economic Research Service, USDA
Washington, DC
roberto.mosheim@usda.gov

and

Robin C. Sickles
Rice University
Houston, Texas
rsickles@rice.edu

March 31, 2020

Abstract

This study explores the spatial effects in nitrogen (N) and phosphorus (P) pollution and drinking water production patterns in agriculture. Two important examples are that water utilities that deliver and treat drinking water in agricultural areas have to deal with excess nitrogen and phosphorus released to the environment by crop and livestock operations, an externality created by the agricultural sector; and, second, that the drinking water production sector in rural areas is a highly fragmented with a multitude of enterprise sizes, organization forms and network densities that have spatial components. In our analysis we present measures of N and P pollution. We employ information collected in section 303(d) of the Clean Water Act: count of impaired water bodies by N/P, and count of point source N/P pollution at the Hydrologic Unit Code 8 (HUC) or sub-basin level and estimate how these variables affect drinking water utilities scale economies, productive efficiency, and scale and scope economies.

Keywords: Drinking Water Utilities; Pollution; Efficiency, Scale, Scope Economies, Spatial Effects.

JEL Q57; Q53; D24

[1] The findings and conclusions in this article are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

1. Introduction

Drinking water pollution has been a top environmental concern since at least 1989 according to Gallup polls [Kaiser and Shapiro (2019)] and has recently been gaining increased traction in policy discussions. In 2018, American Water Works Association (AWWA) CEO David LaFrance applauded the U.S. Congress for passing a Farm Bill that recognizes the importance of protecting drinking water sources from nutrient runoff and that allocated $4 billion dollars over the next ten years to conservation practices that protect sources of drinking water, [AWWA (2020)]. In early 2020, a related important policy change was the U.S. Environmental Protection Agency (EPA) finalizing the repeal of the Waters of the U.S. Rule (WOTUS) which had prohibited the dumping of certain industrial and agricultural pollutants into some 234,000 miles of small streams that provide tap water to more than 117 million people. It was replaced with the Navigable Protection Rule, EPA (2020a).

Agricultural chemicals, in particular, have been a "growing source of environmental problems," including drinking water source pollution affecting millions of people [Snider January 23, 2020]. The focus of the Clean Water Act, however, has been on wastewater treatment and point-source emissions rather than the non-point sources prevalent in agriculture. In an example of the impact of non-point sources used in, Mosheim and Ribaudo (2017), downstream effects played a role when the Des Moines Water Works sued three northwest Iowa counties in federal court for channeling excess nitrogen through an elaborate drainage system into the Raccoon River, the primary source of drinking water for half a million central Iowans, and thus burdening municipal water utility with an alleged cost of about half a million dollars in the 2013-2014 winter, [(Timothy Meinch, *Des Moines Register* March 16, 2015)]. The

complexity of identifying and analyzing sources of water pollution, illustrated in the Des Moines case where the pollutants originated far from the area covered by the utilities, require sophisticated analytic techniques such as spatial econometrics that capture the role of externalities and are based on well-developed datasets.

The U.S. Environment Protection Agency (EPA) maintains data on water pollution including in rivers and lakes. Price and Heberling (2018) in their review article on the effects of source water quality on drinking water costs find, after reviewing twenty-four selected studies on the topic, that the main obstacle to conducting research on water quality in the United States is "data availability," making it difficult to quantify the links between treatment costs and water quality. A similar point is also noted by Keiser and Shapiro (2019). In our study, we employ AWWA data from 2015 and 2016 Water and Wastewater surveys to construct a panel of two outputs (drinking water, wastewater), three inputs (labor, capital and other inputs), and variables that influence the performance of the water utility but are under of the control of the manager (proportion of ground water used in water production), or out of the control of manager (number of impaired water bodies by nitrogen and phosphorus in the hydrologic unit where the water utility is located and number point sources of nitrogen or phosphorus release along the Hydrologic Unit Code 8 (HUC).

Thanks to this dataset, we are able to measure the impact of nitrogen and phosphorus pollution on technical efficiency, scale economies, and economies of scope which ultimately affects costs. We estimate input and output distance functions (which need less data than, for example, cost functions) to predict environmentally sensitive measures of efficiency, calculate scale and scope economies, and spatial effects of nutrient pollution. Given both the importance

of spatial econometrics to the analysis of the externality problem and that that analysis employs a panel, we draw heavily on Sickles and Zelenyuk, (2019) and Baltagi (2015).

In what follows we first explore the main sources of data for our empirical study and describe the variables employed in the analysis. We next present the spatial distance function models. Our empirical findings follow. Finally, we summarize our results and conclusions.

2. Variable construction and descriptive statistics

(insert table 1 here)

Table one describes the variables we employed in the analysis described in section 4. The main source of data employed in this study is the American Water Works Association (AWWA) Rate Survey for 2015 [AWWA (2016b)] and 2016 [AWWA (2016d)], both in computer compact disk. We assembled a balanced panel of water utilities that had observed wastewater and drinking water operations in both years. Water utilities that had data for only one year or that only treated water or produced only water were not employed. The resulting panel covers 59 mostly public firms that had complete information on all the variables representing 42 states as the original data set observed in 2015 and 2016 (Alaska and U.S. territories excluded because of missing observations on water quality), for a total of 114 observations. The combined panel will be referred as AWWA2017.

Water production ranges from a minimum of 1.353 to a maximum of 240 million gallons per day (MGD), and wastewater treatment ranges from 0.49 to 160.46 MGD. The overwhelming majority of the observations are municipally owned. The original data represent 42 states that serve anywhere from 1000 to 5 million customers. The data that we employ in the analysis had as many states represented. We employed the water and wastewater variables from the

AWWA2017 in (MGD) units to construct the drinking water and treated water variables, y_1 and y_2 respectively. Labor (x_1) refers to staff-combined variable (Water and Wastewater Employment). We define capital (x_2) as the variable Assets: Equipment plus Assets Net Plant in AWWA2017 The capital variable is deflated by the PPI industry, BLS (2019). The Other Inputs (x_3) variable is defined as total operations and maintenance costs (OM) minus labor costs. The procedure we employed took account first of the variation of OM costs per account by water utility size and region available in AWWA2017. We also drew on AWWA (2016) "Benchmarking and Performance Indicators" and the 2015 AWWA (2015) and 2016 AWWA (2016c) Compensation Surveys for this task. Regional variation in salaries were assigned using BEA Price Parities by state for the years 2015 and 2016 [BEA (2019)].

The variable *Metro* identifies whether or not the water utility is located in an urban or rural/suburban county. We employ the 2013 Rural-Urban Continuum Codes published by the Economic Research Service. These codes form a classification scheme that distinguishes nonmetropolitan counties by degree of urbanization and adjacency to metro areas, (ERS-USDA (2020). *Size* is a dummy variable that takes the value of 1 if the water utility that has more than 100,000 customers, and 0 otherwise. The variable *Ground* is the total quantity of ground water in MGD going to drinking water production.

We used information from the watershed boundary dataset which maps the water drainage of the United States using a hierarchical system of hydrological unit code (HUC) at different levels of detail. We used various water pollution variables by level hierarchy HUC-8. HUC 8 maps at the sub-basin level are analogous to medium sized river basins (about 2200 nationwide). The HUC classification system represents the coordinated efforts among the United

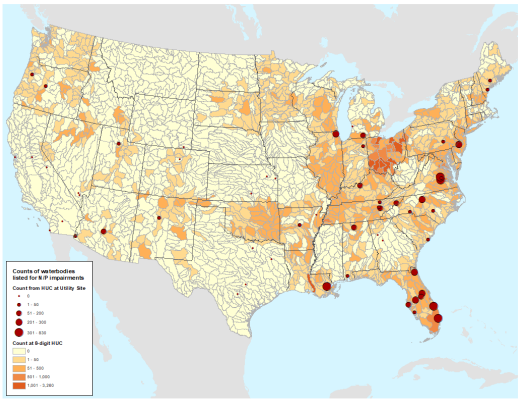
States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), the United States Geological Survey (USGS), and the Environmental Protection Agency (EPA) (USGS (2019)).

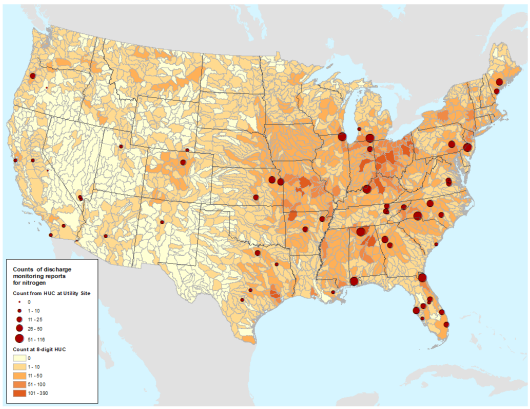
We employ two variables that characterize the water quality at the HUC -8 where the water utility is located. According to EPA, the Clean Water Act requires states, territories and authorized tribes to monitor water pollution and report to EPA every two years on the waters they have evaluated. This process is called assessment and includes deciding which waters do not meet water quality standards because they are too polluted. These degraded waters are called impaired (polluted enough to require action) and are placed on a State list for future actions to reduce pollution. The characteristics by HUC-8 are summarized by EPA (2020).

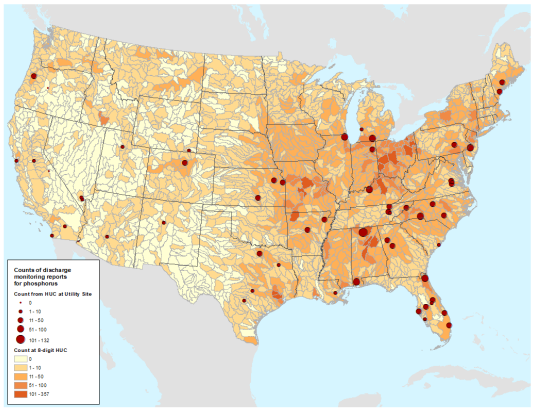
We employ data on EPA ATTAINS-- an acronym for the Assessment, Total Maximum Daily Load (TMDL) Tracking and Implementation System—which is an online system for accessing information about the conditions in the nation's surface waters. Specifically, we examine waterways deemed unsound by Nitrogen or Phosphorus pollution. We also examine point sources permitted to discharge nitrogen or phosphorus into the environment. The National Pollutant Discharge Elimination System (NPDES) keeps track of the number of permits in a given hydrological unit, see EPA (2019b)].

The variables *ImpairedNP, DischargeN,* DischargeP show HUC characteristics at the water utility site: the first variable shows the count of impaired water bodies by Nitrogen and Phosphorus, the second and third, the number of discharge monitoring reports (DMR) for nitrogen, indicating permitted point source pollution for these pollutants, and, third, the number of DMRs for phosphorus. The next three charts also show counts for these three variables for the

United States generally. We also include dummy variables to control for certain utilities facing greater levels of pollution in their source water than average.







3. Methodology

We next discuss how we can represent the technology of a water utility. The water utility is assumed to use the technology

$$T^t = \{(x^t, y^t): x^t \in \mathbb{R}_+^n, y^t \in \mathbb{R}_+^m, x^t \text{ can produce } y^t\} \quad (1)$$

where $x^t = (x_1^t, \ldots, x_n^t)$ and $y^t = (y_1^t, \ldots, y_m^t)$ are the input and output vectors at time $t = 1,2,\ldots,T$.

Let $P^t(x^t)$ be the set of feasible output vectors y^t that are obtainable from each input vector x^t :

$$P^t(x^t) = \{y^t : (x^t, y^t) \in T^t\}. \quad (2)$$

The output distance function [Shephard (1970)] is defined:

$$D_o^t(x^t, y^t) = \min_{\theta} \left\{\theta > 0 : \left(\frac{y^t}{\theta}\right) \in P^t(x^t)\right\}. \quad (3)$$

Given an input vector x^t, the value of $D_o^t(x^t, y^t)$ is equal to 1 if the output vector lies in the boundary of $P^t(x^t)$. Modelling the effect of time as an exogenous variable the output distance function can be rewritten as $D_o(x, y, t)$.

For each output vector y^t, let $L^t(y^t)$ be the set of feasible input vectors x^t that can be obtained from each output vector y^t:

$$L^t(y^t) = \{x^t : (y^t, x^t) \in T^t\}. \quad (4)$$

The input distance function (Shephard (1970)) is:

$$D_I^t(x^t, y^t) = \frac{\max}{\theta} \left\{\theta > 0 : \left(\frac{x^t}{\theta}\right) \in L(y^t)\right\}. \quad (5)$$

Given an input vector x^t, the value of $D_I^t(x^t, y^t)$ is equal to 1 if the input vector lies in the boundary of $L^t(y^t)$.

4. Empirical Results

For the specification and estimation of the input and output distance functions we followed several well received articles in their empirical application: Kumbhakar, et al. (2007); Coelli and Perlman (2000), Nemoto and Furumatsu (2014), Feng and Serletis (2009), Inanoglu et

al. (2015), Lovell et al. (1994). We also employed spatial regression techniques discussed in Elhorst (2014), Arbia (2014), and Drukker et al. (2013).

Kumbhakar et al. discuss parametric input and output frontier for a cross section. Important aspects are that inefficiency is non-negative if the distance function is output oriented whereas it would be non-positive if the distance function is input oriented. We also impose the homogeneity of degree one in outputs by dividing water production by the water treatment variable in the case of the output distance function. restrictions by dividing outputs. In the case of the input distance function we imposed homogeneity of degree on inputs by dividing labor and capital by other inputs.

We employ panel data for our models. Our input-oriented model is:

$$-\ln(O)_{it} = \beta_o + \sum_{j=1}^{M=2} \gamma_j \ln y_{j,it} + \frac{1}{2}\sum_{j=1}^{M=2}\sum_{k=1}^{M=2} \gamma_{jk} \ln y_{j,it} \ln y_{k,it} + \sum_{k=1}^{N=K-1=2} \beta_k \ln\left(x_k/O\right)_{it} + \frac{1}{2}\sum_{k=1}^{K-1=2}\sum_{l=1}^{L-1=2} \beta_{kl} \ln\left(x_k/O\right)_{it}\left(x_l/O\right)_{it} + \sum_{j=1}^{M=2}\sum_{k=1}^{K-1=2} \delta_{jk} \ln y_{j,it} \ln\left(x_k/O\right)_{it}. \quad (6)$$

We tested various specifications of equation (8) and found the quadratic input terms insignificant. Consequently, our final specification is Translog in outputs and Cobb-Douglas in inputs following Sickles, Good and Getachew (2002). Moreover, our final specification incorporates the variable y2016 that measures technical change relative to 2016:

$$\ln(D_I)_{it} = -\ln(O)_{it} = \beta_o + \sum_{j=1}^{m=2} \alpha_j \ln y_{j,it} + \frac{1}{2}\sum_{j=1}^{m=2}\sum_{k=1}^{n=2} \alpha_{ij} \ln y_{j,it} \ln y_{k,it} + \sum_{k=1}^{n=2} \beta_k \ln\left(x_k/O\right)_{it} + \delta_t Y2016. \quad (7)$$

This model is then transformed into a spatial autoregressive (SARAR (1,1)) specification following Lee and Yu (2010a), Lee and Yu (2010b) and Arbia (2014):

$$\ln(D_I)_{it} = \lambda W \ln(D_I)_{it} + \sum_{j=1}^{m=2} \alpha_m \ln y_{m,it} + \frac{1}{2} \sum_{j=1}^{m=2} \sum_{k=1}^{n=2} \alpha_{jk} \ln y_{j,it} \ln y_{k,it} + \sum_{k=1}^{n=2} \beta_k \ln\left(x_k / O\right)_{it} + \delta_1 y2016 + c_n$$

$$+u_{it}, \quad |\lambda| < 1,$$

$$u_{it} = \rho W u_{it} + v_{it} \qquad |\rho| < 1. \tag{8}$$

Arbia (2014) refers to the case where $\lambda \neq 0, \rho \neq 0$ using the SARAR (1, 1) acronym for Spatial Autoregressive with additional Autoregressive error structure (Kelejian and Prucha (1998)). We row-standardized *W,* and given that $|\lambda| < 1$ and $|\rho| < 1$ insures a solution to the parameters of equation (10). We followed Arbia (2014) and STATA (2017) when we specified our spatial autoregressive models for panel data in equation (10) above. Here $\ln(D_I)_{it} = (d_{I1t}, d_{I2t}, ..., d_{Int})'$ is an $n \times 1$ vector of observations for n panels at the time period *t*; W is an $n \times n$ spatial weighting matrix exogenously given and row-standardized; $\ln(y)_{it} = (y_{1t}, y_{2t}, ..., y_{nt})$ is an $n \times 1$ vector of observations for *n* panels of observations for drinking water and treated water; $\beta_j W \left(\frac{x}{O}\right)_{it} + \gamma_j \left(\frac{\overline{x}}{O}\right)_{it}$ is an $n \times 1$ vector of observations for *n* panels of observations for labor-other inputs water and capital-other input ratios multiplied or not by *W*; the variable z_1 is a dummy variable equal to one the number of impaired of water bodies is above average for the water utilities and 0 if it is below average; u_{it} is an $n \times 1$ vector of spatially lagged errors; v_{it} is an *i.i.d.* across *i* and *t* with variance; c_n is random effects with mean 0 and variance σ_c^2; ρ is the spatial dependence parameter; and $\lambda, \alpha, \beta, \gamma, \rho, \delta$ are parameters to be estimated.

The output distance is specified as:

$$-\ln(y_2)_{it} = \beta_o + \alpha_1 \ln\left(y_1/y_2\right)_{it} + \frac{1}{2}\alpha_{11} \ln\left(y_1/y_2\right)^2_{it} + \sum_{k=1}^{K} \beta_k \ln(x_k)_{it}$$
$$+\frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{L} \beta_{kl} \ln(x_k)_{it} \ln(x_l)_{it} + \sum_{k=1}^{K} \delta_{1k} \ln\left(y_1/y_2\right)_{it} \ln(x_k)_{it} + \delta_t y2016. \quad (9)$$

As the model above does not support the quadratic input terms we adjust the specification to:

$$\ln(D_O)_{it} = -\ln(y_2) = \beta_o + \alpha_1 \ln(y_1/y_2) + \frac{1}{2}\alpha_{11} \ln(y_1/y_2)^2 + \sum_{k=1}^{K} \beta_k \ln(x_k)_{it} + \delta_t y2016. \quad (10)$$

This model is then transformed into a spatial autoregressive (SARAR (1,1)) specification:

$$\ln(D_O)_{it} = \lambda W \ln(D_O)_{it} + \alpha_1 \ln\left(y_1/y_2\right)_{it} + \frac{1}{2}\alpha_{11} \ln\left(y_1/y_2\right)^2_{it} + \sum_{k=1}^{K} \beta_k \ln x_{it} + \delta_t y2016 + c_n + u_{it},$$

$$|\lambda| < 1, \quad (11)$$

$$u_{it} = \rho W u_{it} + v_{it}, \quad |\rho| < 1.$$

We estimated four distance functions: two input-oriented and two output-oriented specifications. Of these four models two are spatial and two nonspatial (regression models.) The non-spatial models incorporate the same variables used in the spatial model. We estimated both the input and output distance spatial functions by correlated random effects. Spatial correlated effects[2] were preferred to random spatial effects models. We present the two econometric estimation results of the spatial input and output distance functions in tables 3a and 3b, and regular regression counterparts in Table3areg and Table3breg. A spatial model is better than a

[2] See Wooldridge (2013)

non-spatial one by the significance of the estimated value of the spatial autocorrelation parameter, $\widehat{\rho}$ (STATA SP, p53). Also, model selection statistics (AIC and BIC) identify first that the spatial models as preferred to the non-spatial models, and second that the spatial input distance is preferred to the output distance function. In what follows we discuss both the spatial input and output distance function models calculation of scale economies just to compare the scale economy results and confirm our exclusion of model 2. Then we proceed to estimate scope economies just using model 1.

Tables 3a and 3b shows the estimates for the input and output distance functions. The implied elasticities for the input distance function for outputs 1 and 2 are $\partial \ln D_I / \partial \ln y_1 = -0.637$ and $\partial \ln D_I / \partial \ln y_2 = -0.159$. The bootstrap standard errors are respectively 0.005 and 0.022. The input function equation for output 1 is

$$\partial \ln D_I / \partial \ln y_1 = \beta_1 + 2\beta_3 \ln y_1 + \beta_5 \ln y_2 , \tag{12}$$

and output 2 is

$$\partial \ln D_I / \partial \ln y_2 = \beta_2 + 2\beta_4 \ln y_2 + \beta_5 \ln y_1 . \tag{13}$$

The implied elasticity of the output distance with respect to water production and water treatment are: $\partial \ln D_o / \partial \ln y_1 = 3.589$ and $\partial \ln D_o / \partial \ln y_2 = 3.827$.

The output distance function equation for output 1 is

$$\partial \ln D_o / \partial \ln y_1 = \beta_1 1/(y_1/y_2)/(1/y_2) + \beta_3 (\ln y_1 y_2)(1/y_2) \tag{14}$$

and for output 2,

$$\partial \ln D_o \big/ \partial \ln y_2 = \beta_1 1/(y_1/y_2)/(1/y_1) + \beta_3(\ln y_1 y_2)(1/y_1). \quad (15)$$

The bootstrap standard errors are respectively 0.587 and 0.651.

Also, Table 3a shows that all the input elasticities are negative indicating the estimated input distance function meets monotonicity properties: the inputs decrease and outputs increase with respect to distance. Homogeneity of degree one has been imposed by dividing distance and input quantities by the input variable x_3, other inputs. Table 3b shows the output distance function results. Homogeneity of degree one in outputs was imposed by dividing water production output by water treatment output ($y_1 \big/ y_2$). The estimated output function meets the monotonicity properties: increasing in inputs and decreasing in outputs.

The y2016 variables have different signs whether it is incorporated in an input or output orientated distance function in as expected; both functions point to an increase in technical change of 2016 relative to 2015. Also, both input and output distance functions show that a greater share of ground water sourced by the water utility is predicted to increase water utility inefficiency and hence costs. The variable ImpairedNP is a dummy variable that represents the number of water bodies impaired by N and P in the HUC where the water utility operates. If the number of impaired water bodies is greater than the average in the HUC, the dummy variable is equal to one. Otherwise, it is equal to zero. The estimate in the output distance is highly significant.

The variables DischargeN and DischargeP captures the number of point pollution sources within the HUC where the water utility operates. If the DischargeN and DischargeP dummy variables are equal to 1 it means that the water utility operates in a HUC where the number of point sources is greater than average, zero otherwise. If the effect of DischargeN is positive it

means that water utilities facing larger than average number of point sources in their HUC face increase use of inputs or decrease production of outputs. The process by which N affects water utility cost might be more rapid or more complex than for P and that might the different signs of the effects. The effects of the DischargeN and DischargeP are highly significant for both input or output distance functions and have opposite signs as expected.

After estimating input-orienting inefficiency we predicted inefficiency. We exponentiated the result and then multiplied by the other input variable. We then divided the result by the minimum calculation of this result for the 114 observations. Similarly, after we predicted output-oriented inefficiency we exponentiated the result before multiplying it by y_2. We then divided this prediction by the maximum output result for this calculation from the 114 observations. Table 4 shows the distance results for the four models. As can be seen from this table results for models 1 and 3 are very similar and results 2 and 4 diverge from these models. The discussion in the previous paragraph would make the inefficiency results from models 1 and 3 more robust than those from model 2 and 4.

We calculated input oriented and output-oriented scale economies using two tray returns to scale formulas. The first is Input-Oriented [Nemoto and Furumatsu [(2014, p.8)]:

$$RTS_I = -\frac{1}{\sum_{s=1}^{s=2} \partial \ln D_I(x,y,t) \Big/ \partial \ln y_s}. \tag{16}$$

The second is Output-Oriented [Feng and Serletis (2009, p. 11)]:

$$RTS_O = -\sum_{s=1}^{s=3} \partial \ln D_O(x,y,t) \Big/ \partial \ln x_s. \tag{17}$$

The pattern of Table 4 is the same as those of table 2: results for models two and four diverge and results for models 1 and 3 do not—they are a great deal more robust. We use the delta method to calculate the standard errors for the scale economies. We simply use model 1 to calculate scope economies the input distance function.

Economies of scope between outputs *i* and *j* is calculated by the derivatives of either the input distance or output distance functions (Hajargashat, et al., 2006; Inanoglu, et al., 2015):

$$C_{yy}/C = D_y D_y' - D_{yy} + D_{yx}\left[D_{xx} + D_x D_x'\right]^{-1} D_{xy}. \quad (18)$$

We restricted equation (18) so that it exhibits input homotheticity so that equation (18) becomes:[3]

$$C_{yy}/C = 2D_y D_y' - D_{yy}. \quad (19)$$

Equation (19) has a positive sign which Hajargashat, et al. (2008, p.186) associate with the existence of scope economies.[4] This result indicates that water utilities that separate water production and treatment facilities will have higher costs than the ones that do not. We calculated scope economy averages by exogenous variables employed in the analysis. We underscore two results: first, scope economies for rural utilities is 0.201 and for urban ones are 0.302; second, water utilities that face higher than average environmental pollution in their source water have lower economies of scope. These economies of scope are: DischargeN, 0.276, DischargeP,

[3] Without this restriction the data cannot identify a reasonable technology. This restriction is consistent with a general technology, more so than the standard Cobb-Douglas technology.

[4] The calculation of equation (19) results in four quadrants of size (114×114). The upper right quadrant and the lower left quadrant involve cross derivatives between the two outputs. We constrained these cross derivatives to be equal. We calculated the analytical derivatives of the distance functions following Hajargashat, et al. (2006, p.17). The diagonal of the upper right matrix is the implementation of (19).

0.280, and ImpairedNP, 0.291. Otherwise, scope economies are 0.303, 0.302 and 0.291 for these variables respectively.

The total impact of an independent variable x is the average of the marginal effects it has on the reduced-form mean:

$$\frac{1}{n^2T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial E(D_{Iit} \mid X_{it}, W; Y_{it})}{\partial x_{jt}}. \tag{20}$$

The direct impact of an independent variable x is the average of the direct, or own, marginal effects:

$$\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial E(D_{Iit} \mid X_{it}, W; Y_{it})}{\partial x_{it}}. \tag{21}$$

The indirect impacts of an independent variable x

$$\frac{1}{nT(n-1)}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\frac{\partial E(D_{Iit} \mid X_{it}, W; Y_{it})}{\partial x_{it}}. \tag{22}$$

A further discussion of these measures of spillovers can be found in Le Sage and Pace (2009, 36-37) and STATA (2017, pp 218). Results are displayed in Table 5. The most important aspect to note from these results is the strength of the indirect effects of point sources of N and P pollution on water utility performance which underscores the externality motivation of the spatial analysis we applied. The strength of the Ground variable could underscore the interconnection between water utilities.

5. Summary of Results and Conclusions

This paper models the environmental effects of nutrients (nitrogen and phosphorus) coming from agriculture on drinking water production. We employed a dataset drawn from

myriad sources to analyze inputs, outputs, source water pollution variables and other exogenous variables affecting water utilities. We estimated input and output distance functions to derive environmentally sensitive measures of efficiency (input- and output-oriented), scale economies, scope economies and spatial effects. We find that input-oriented results are more robust than output-oriented ones. Spatial model specification tests indicated application of spatial econometric models is more appropriate than regular regression. In our analysis we present measures of nitrogen and Phosphorus pollution: ImpairedNP, number of impaired water bodies impaired by Nitrogen or phosphorus. DischargeN and Discharge P number of facilities with permits to discharge N or P in the utility. These variables constitute measures of N and P point and non-point pollution in the hydrological unit where the utility captures its water to produce drinking water. We showed that these variables do indeed affect scale, scope and efficiency of the water utilities.

We take advantage of the ability of spatial econometrics to model externalities. We find large indirect effects of point and nonpoint sources of nitrogen and phosphorus pollution. We find that the sector exhibits significant scope economies between drinking water production and drinking water quality and increasing returns to scale. A water utility might face the choice of how much water quality to maintain to produce a given level of safe drinking water. However, there are other factors outside the control of the manager that might affect the costs of operation. We showed that the presence of certain contaminants in the source water that the drinking water utility employ affects a myriad of production measures which translate ultimately into increased costs. We hope that this analysis might be of use to both local and national policy makers concerned with the economic impact of environmental regulations.

Compliance with Ethical Standards

Conflict of Interest: Roberto Mosheim declares that he has no conflict of interest. Robin Sickles declares that he has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

6. References

Arbia, G. (2014), A Primer for Spatial Econometrics With Applications in R. London: Palgrave Macmillan.

AWWA (2015), 2015 Compensation Survey (Small, Medium and Large), Denver: CO: American Water Works Association.

AWWA (2016), "Benchmarking and Performance Indicators for Water and Wastewater: 2016 Edition: Survey Data and Analyses Report" Denver, CO: American Water Works Association.

AWWA (2016b), 2015 Water and Wastewater Rate Survey Update, Interactive Database, CD, Denver: CO: American Water Works Association and Raftelis Financial Consultants.

AWWA (2016c) 2016 Compensation Survey (Small, Medium and Large), Denver: CO: American Water Works Association.

AWWA (2016d), 2016 Water and Wastewater Rate Survey Update, Interactive Database, CD, Denver: CO: American Water Works Association and Raftelis Financial Consultants.

AWWA (2020), AWWA Articles 2018, AWWA commends Congress for including drinking water protection measures in Farm Bill, December 12, 2018, https://www.awwa.org/AWWA-Articles/awwa-commends-congress-for-including-drinking-water-protection-measures-in-farm-bill [accessed January, 29, 2020].

Bureau of Economic Analysis [BEA (2019)], Regional Price Parities by State and Metro Area, years 2015 and 2016. https://www.bea.gov/data/prices-inflation/regional-price-parities-state-and-metro-area, accessed 1/29/2019).

Coelli, T, and S. Perelman (2000), "Technical Efficiency of European Railways: A Distance Function Approach" Applied Economics, 32, 1967-1976.

Drukker, D. , H. Peng, I. Prucha, R. Raciborski (2013), Creating and managing spatial-weighting matrices with the spmat command , The Stata Journal 13, Number 2, pp. 242–286.

ERS-USDA, Rural-Urban Continuum Codes,
(2020), https://www.ers.usda.gov/data-products/#!topicid=14838&subtopicid=14910, [accessed March 4, 2020].

Elhorst, J Paul (2014), Spatial Econometrics From Cross-Sectional Data to Spatial Panels. Springer Briefs in Regional Science.

Guohua F. and A. Serletis (2009), Efficiency, Technical Change, and Returns to Scale in Large U.S. Banks: Panel Data Evidence from an Output Distance Function Satisfying Theoretical Regularity, Working Paper 5/09, Department of Econometrics and Business Statistics, Monash University, Australia.

Hajargashat, G., T. Coelli and D.S.P. Rao (2006), A Dual Measure of Economies of Scope, Centre for Efficiency and Productivity Analysis, Working Paper Series No. 03/2006. School of Economics, University of Queensland.

Hajargashat, G., T. Coelli and D.S.P. Rao (2008), A Dual Measure of Economies of Scope, *Economic Letters*, 100: p185-188

Inanoglu, H., M. Jacobs, J. Liu and R. Sickles (2015), Analyzing Bank Efficiency: Are "too-big-to-fail" Banks Efficient?", McMillan Palgrave Handbook of post crisis financial modeling.

Keiser, D., and J. Shapiro (2019), "US Water Pollution Regulation over the Past Half Century: Burning Waters to Crystal Springs?" Journal of Economic Perspectives 33(4): 51-75.

Kelejian, H.H. and I.R. Prucha (1998), "A Generalized Spatial Two-Stage Least Squares Procedures for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." Journal of Real Estate Finance and Economics, 17, 1998, 99-121.

Kumbhakar, S. C., L. Orea A. Rodríguez E. Tsionas, (2007), "Do we estimate an input or an output distance function? An application of the mixture approach to European railways." Journal of Productivity Analysis 27:87–100.

LeSage J. and R. Pace (2009), "Introduction to Spatial Econometrics.", 1st Edition, CRC Press.

Lee and Yu (2010a), "Estimation of spatial autoregressive panel data models with fixed effects" Journal of Econometrics 154: 165-185.

Lee and Yu (2010b), "Some recent developments in spatial panel data models" Regional Science and Urban Economics 40(5): 255-271.

Lovell C.A.K., Travers P., Richardson S., Wood L. (1994) Resources and Functionings: A New View of Inequality in Australia. In: Eichhorn W. (eds) Models and Measurement of Welfare and Inequality. Springer, Berlin, Heidelberg.

Meinch, Timothy, "Water Works requests damages in federal suit." Des Moines Register, March 16,2015.

Mosheim, R. and M. Ribaudo (2017), "Costs of Nitrogen Runoff for Rural Water Utilities: A Shadow Cost Approach," Land Economics 93(1): 12-39.

Nemoto, J. and N. Furumatsu (2014), "Scale and Scope Economies of Japanese Private Universities Revisited with an Input Distance Function Approach," Manuscript, NYU- Stern

Oxford Handbook of Panel Data (2015), Edited by Badi Baltagi, New York: Oxford University Press.

Price, J. I. & Heberling, M. T., (2018), "The Effects of Source Water Quality on Drinking Water Treatment Costs: A Review and Synthesis of Empirical Literature," Ecological Economics, vol. 151(C), pages 195-209.

Shepard, R.W. (1970) Theory of Cost and Production Function. Princeton University Press, Princeton.

Sickles, R., D. Good and L. Getachew, (2002). "Specification of Distance Functions Using Semi- and Nonparametric Methods with an Application to the Dynamic Performance of Eastern and Western European Air Carriers," Journal of Productivity Analysis, vol. 17(1), pages 133-155.

Sickles, R. and V. Zelenyuk, (2019). Measurement of Productivity and Efficiency: Theory and Practice. Cambridge: Cambridge University Press.

Snider, A., "Trump erodes water protections: 6 things to know." Politico, January 23, 2020.

STATA (2017), STATA Spatial Autoregressive Models Reference Manual, Release 15. STATA: College Station, TX.

U.S. Bureau of Labor Statistics (BLS) (2019), PPI Industry Data, Series Id, PCUOMFG--OMFG—(https://data.bls.gov/timeseries/PCUOMFG--OMFG--, accessed 1/29/2020)

United States Environmental Protection Agency (2020), Navigable Waters Protection Rule, https://www.epa.gov/nwpr/about-waters-united-states [accessed January, 29, 2020].

United States Environmental Protection Agency (2020), The characteristics by HUC-8 are summarized in https://gispub2.epa.gov/NPDAT/DataDownloads.html. [accessed 1/29/2020].

United States Geological Survey (2019), https://water.usgs.gov/GIS/huc.html [Accessed 12/30/2019].

United States Environmental Protection Agency (2019), https://gispub2.epa.gov/NPDAT/DataDownloads.html.

United States Environmental Protection Agency (2019b), see https://www.epa.gov/npdes.

Wooldridge, J. (2013), "INTRODUCTION AND LINEAR MODELS: Correlated Random Effects Panel Data Models" IZA Summer School in Labor Economics", May 13-19 2013.

Table 1.- Summary Statistics Spatial Distance Functions Variables

Variable (Units)	Mean	Std. Dev.	Min	Max
Latitude (Decimal (6))	35.66	4.80	26.26	44.64
Longitude (Decimal (6))	-92.61	15.06	-123.11	-69.77
y_1 = Drinking Water (MGD)	36.79	49.20	1.35	240
y_2 = Treated Water (MGD)	26.32	33.65	0.49	160.46
x_1 = Labor (Staff)	297.04	453.74	5.00	2158
x_2 = Capital (Million $)	405.00	657.00	0.39	4620
x_3 = Other Inputs (Million $)	19.00	20.80	0.61	151
Metro (Dummy: urban=1, rural =0)	0.93	0.26	0	1
Y2016 (Dummy: y2016=1, if year=2016, else 0)	0.50	0.50	0	1
Ground (Total source ground water (MG))	92,44	671,98	0	6.82M
ImpairedNP (Dummy: N/P Imp. Waters in HUC)	0.21	0.41	0	1
DischargeN, (Dummy: Facilities permits N in HUC)	0.28	0.45	0	1
DischargeP, (Dummy: Facilities permit P in HUC)	0.33	0.47	0	1
Observations (i = 57, t=2 2015, 2016)		114		

Table 2. – Weight Matrix (Distance), *W*

Weight Element	Weight (Rounded)	Miles
Minimum > 0	0.00065	1548.22
Mean	0.00824	121.37
Max	0.69390	1.44

Table 3a_reg. Regression estimates of input-oriented distance function

(standard errors in parentheses)

Variable	Par.	Estimate	Variable	Par.	Estimate
Input Distance Function, Regression					
$\ln(y_1)$	α_1	-0.570***	$\overline{x_1/x_3}$	$\overline{\beta_{23}}$	-0.006***
		(0.165)			(0.002)
$\ln(y_2)$	α_2	0.141	$Y2016$	δ_t	-0.271***
		(0.151)			(0.076)
$[\ln(y_1)]^2$	α_{11}	0.028	Constant	β_o	-13.08***
		(0.067)			(0.892)
$[\ln(y_2)]^2$	α_{22}	-0.006			
		(0.055)			
$\ln(y_1)\times\ln(y_2)$	α_{12}	-0.093			
		(0.113)			
$\ln\left(x_1/x_3\right)$	β_{13}	0.127**			
		(0.070)			
$\overline{x_1/x_3}$	$\overline{\beta_{13}}$	25290***			
		(5,544)			
$\ln\left(x_2/x_3\right)$	β_{23}	0.071			
		(0.044)			
Ground	γ_1	-2.59e-08***	Metro	γ_5	-0.141
		(5.95e-08)			(0.185)
DischargeN	γ_2	-0.122			
		(0.228)			

DischargeP	γ_3	-0.056 (0.232)
ImpairedNP	γ_4	-0.0005* (0.0003)

Note: * P = <0.10 ; ** P = <0.05 ; *** P = <0.01

Variance Parameters

$ADJR^2$ 0.879

Model Selection Statistics (degrees of freedom in parentheses)

Log Likelihood -64.02

Penalized Likelihood Criteria d.f. (14)

Akaike information (AIC)= 156.043

Bayes Information (BIC)= 194.345

F(13,100) =64.11 Prob > F=0.0000

Table 3b_reg. Regression estimates of output-oriented distance function

(standard errors in parentheses)

Variable	Par.	Estimate	Variable	Par.	Estimate
Output Distance Function, Regression					
$\ln(y_1)$	α_1	-0.570***	$\overline{x_1/x_3}$	$\overline{\beta_{23}}$	-0.006***
		(0.165)			(0.002)
$\ln(y_2)$	α_2	0.141	$Y2016$	δ_t	-0.271***
		(0.151)			(0.076)
$[\ln(y_1)]^2$	α_{11}	0.028	Constant	β_o	-13.08***
		(0.067)			(0.892)
$[\ln(y_2)]^2$	α_{22}	-0.006			
		(0.055)			
$\ln(y_1)\times\ln(y_2)$	α_{12}	-0.093			
		(0.113)			
$\ln\left(x_1/x_3\right)$	β_{13}	0.127**			
		(0.070)			
$\overline{x_1/x_3}$	$\overline{\beta_{13}}$	25290***			
		(5,544)			
$\ln\left(x_2/x_3\right)$	β_{23}	0.071			
		(0.044)			
Ground	γ_1	-2.59e-08***	Metro	γ_5	-0.141
		(5.95e-08)			(0.185)
DischargeN	γ_2	-0.122			
		(0.228)			

DischargeP	γ_3	-0.056 (0.232)
ImpairedNP	γ_4	-0.0005* (0.0003)

Note: * P = <0.10 ; ** P = <0.05 ; *** P = <0.01

Variance Parameters

$ADJR^2$ 0.880

Model Selection Statistics (degrees of freedom in parentheses)

Log Likelihood -39.55

Penalized Likelihood Criteria d.f. (16)

Akaike information (AIC)= 111.092

Bayes Information (BIC)= 154.871

F(15,98) =56.09 Prob > F=0.0000

Table 3a. Maximum likelihood estimates of input-oriented distance function

(standard errors in parentheses)

Variable	Par.	Estimate	Variable	Par.	Estimate
Input Distance Function, Correlated Random Effects. Dependent variable, $-\ln x_3$					
$\ln(y_1)$	α_1	-0.517***	$\overline{x_1/x_3}$	$\overline{\beta_{23}}$	-0.005***
		(0.125)			(0.002)
$\ln(y_2)$	α_2	0.118	$Y2016$	δ_t	-0.178***
		(0.107)			(0.032)
$[\ln(y_1)]^2$	α_{11}	0.124*	Constant	β_o	-13.383***
		(0.048)			(0.366)
$[\ln(y_2)]^2$	α_{22}	0.107***			
		(0.035)			
$\ln(y_1)\times\ln(y_2)$	α_{12}	-0.303***			
		(0.074)			
$\ln(x_1/x_3)$	β_{13}	0.140***			
		(0.025)			
$\overline{x_1/x_3}$	$\overline{\beta_{13}}$	23,534***			
		(4,471)			
$\ln(x_2/x_3)$	β_{23}	0.110***			
		(0.034)			
*Inverse distance weight matrix, w*02					
Ground	γ_1	9.63e-07***	Metro	γ_5	2.049
		(3.30e-07)			(1.502)
DischargeN	γ_2	9.302***			

		(1.608)			
DischargeP	γ_3	-7.485***	lndoi		-0.140
		(1.245)			(0.098)
ImpairedNP	γ_4	0.0009	e.lndoi	$\widehat{\rho}$	0.566**
		(0.0008)			(0.274)

Note: * P = <0.10 ; ** P = <0.05 ; *** P = <0.01

Variance Parameters

Std. Error	σ_u	0.269
		(0.029)
Std. Error	σ_ε	0.100
		(0.001)

Model Selection Statistics (degrees of freedom in parentheses)

Log Likelihood 21.386

Penalized Likelihood Criteria d.f. (19)

Akaike information (AIC)= -4.77

Bayes Information (BIC)= 47.21

Wald test of spatial terms: $\chi^2(7)$ =49.80 Prob > χ^2 =0.0000

Table 3b. Maximum likelihood preferred estimates of output-oriented distance function

(standard errors in parentheses)

Variable	Par.	Estimate	Variable	Par.	Estimate
Output Distance Function, Correlated Random Effects. Dependent variable, $-\ln y_2$					
$\ln\left(y_1/y_2\right)$	α_{12}	0.608***	$\ln(x_1)$	β_1	-0.181***
		(0.082)			(0.030)
$\overline{y_1/y_2}$	$\overline{\alpha_{12}}$	-0.189	$\ln(x_2)$	β_2	-0.189***
		(0.160)			(0.045)
$\frac{1}{2}\ln\left(y_1/y_2\right)^2$	α_{12}^2	-0.279***	$\ln(x_3)$	β_3	-0.652***
		(0.099)			(0.057)
$\frac{1}{2}\left(\overline{y_1}/\overline{y_2}\right)^2$	$\overline{\alpha_{12}^2}$	0.087***	Y2016	δ_t	-0.156***
		(0.035)			(0.043)
			Constant	β_o	-12.286***
					(0.961)
*Inverse distance weight matrix, w*02					
Ground	γ_1	-1.79e-06***	ImpairedNP	γ_5	-0.004***
		(4.37e-07)			(0.001)
DischargeN	γ_2	-6.020***			
		(2.082)			
DischargeP	γ_3	4.873***	lndoo		-0.366***
		(1.690)			(0.176)
Metro	γ_4	0.140	e.lndoo	$\hat{\rho}$	1.402***
		(0.404)			(0.147)
Note: * P = <0.10 ; ** P = <0.05 ; *** P = <0.01					

Variance Parameters		
Std. Error	σ_u	0.381
		(0.043)
Std. Error	σ_ε	0.112
		(0.012)

Model Selection Statistics (degrees of freedom in parentheses)

Log Likelihood -4.608

Penalized Likelihood Criteria d.f. (18)

Akaike information (AIC)= 45.22

Bayes Information (BIC)= 94.47

Wald test of spatial terms: $\chi^2(7) = 126.26$ Prob > χ^2 =0.0000

Table 4. Predicted Inefficiency and Economies of Scale and Scope

(Mean, Standard Error, Minimum and Maximum)[5]

	Model 1	Model 2	Model 3	Model 4
		Predicted Inefficiency		
Input Oriented	2.301		2.280	
	(0.667)		(0.759)	
	1		1	
	4.758		4.880	
Output Oriented		2.908		3.813
		(1.292)		(1.958)
		1		1
		10.078		14.362
		Scale Economies		
Ray Input	1.271		1.242	
	(0.057)		(0.056)	
Ray Output		0.958		0.949
		(0.002)		(0.001)
		Scope Economies		
Input Oriented	0.295			
	(0.059)			
	0.521			
	0.020			

[5] Models 1 and 2 are spatial and model 3 and 4 are nonspatial.

Table 5. Average Direct, Indirect and Total Effects [a]

(Standard Errors in Parentheses)

VARIABLE	Model 1	Model 2	VARIABLE	Model1	Model2
Metro			DischargeN		
Direct	0.019	-0.003	Direct	0.084	0.127*
	(0.028)	(0.008)		(0.071)	(0.078)
Indirect	1.053	0.056	Indirect	4.782***	-2.390***
	(0.845)	(0.162)		(1.021)	(0.829)
Total	1.072	0.053	Total	4.867***	-2.263***
	(0.872)	(0.154)		(1.079)	(0.796)
ImpairedNP			DischargeP		
Direct	8.42e-06	0.00008	Direct	-0.068	-0.103*
	(0.00001)	(0.00006)		(0.056)	(0.062)
Indirect	0.0005	-0.001***	Indirect	-3.848***	1.934***
	(0.0004)	(0.0005)		(0.770)	(0.679)
Total	0.0005	-0.001***	Total	-3.916***	1.832***
	(0.0004)	(0.0004)		(0.813)	(0.654)
Ground					
Direct	8.78e-09	3.78e-08**			
	6.84e-09	(2.02e-08)			
Indirect	4.95e-07***	-7.11e-07***			
	1.68e-07	(1.80e-07)			
Total	5.04e-07***	-6.74e-07***			
	1.72e-07	(1.77e-07)			

[a] * 10%, ** 5% and *** 1% significance level.