Estimating a Semiparametric Spatial Autoregressive Stochastic Frontier Model

Kien C. Tran* and Mike G. Tsionas†

September 14, 2020

Abstract

This paper proposes a semiparametric spatial autoregressive stochastic frontier model where the functional form of the frontier is modelled nonparametrically. A three-step estimation procedure is considered where in the first two steps, a constrained semiparametric profile GMM is used to obtain the estimates of the spatial parameter and the unknown smooth function of the frontier; whilst in the final step, the remaining parameters of the model can be estimated using ML procedure. We derive the limiting distributions of the proposed estimators for both parametric and nonparametric components in the model. Monte Carlo simulations reveal that our proposed estimators perform well in finite sample.

Keywords:

JEL Classification: C12, C13, C14, C21

*Corresponding author. Department of Economics, University of Lethbridge, Lethbridge, AB, T1K 3M4, Canada; email: kien.tran@uleth.ca

†Lancaster University Management School, Lancaster, LA1 4YX, U.K.; email: m.tsionas@lancaster.ac.uk

1 Introduction

The main idea of this paper is to propose a nonparametric spatial autoregressive (SAR) SF model.

2 The Model

Consider the following semiparametric spatial autoregressive (SAR) stochastic frontier model (SFM)[1]:

$$y_i = \rho \sum_{j \neq i} w_{ij} y_j + m(x_i) + v_i - u_i, \quad i = 1, \ldots, n, \tag{1}$$

where y_i is the (scalar) variable representing log of output of firm i; w_{ij} is the $(i,j)^{th}$ element of a given $(n \times n)$ non-stochastic spatial weighting matrix $\boldsymbol{W}_n$ such that $w_{ii} = 0$ for all i; ρ is an unknown scalar spatial lag parameter, $m(x_i)$ is an unknown smooth function defined on $\mathcal{R}^d$ where x_i is a $(1 \times d)$ vector of log of inputs; v_i is a two-sided symmetric random error, $u_i > 0$ is a one-sided error representing technical inefficiency. Following standard practice, we assume that $v_i \sim i.i.d.\ N(0, \sigma_v^2)$, $u_i \sim i.i.d.\ N^+(0, \sigma_u^2)$. In addition, we assume that x_i are uncorrelated with both v_i and u_i, and v_i and u_i are independent. In this paper, we are interested in estimating $m(\cdot)$ and ρ consistently in such a way that the estimate of $m(\cdot)$ satisfies the requisite axioms of production (or cost), and in using these estimates to predict the firm-specific inefficiency levels which is our primary interest.

Note that, the assumption that the inefficiency term u_i is independent of v_i can be relaxed by introducing a copula function to model the joint distribution of (u_i, v_i) similarly to Amsler et al. (2016). Other assumptions on the marginal distributions of u_i, and/or allow for u_i to depend on a set of environmental variables can also be considered (see for example, Kumbhakar et al. (1991), Huang and Liu (1994), Battese and Coelli (1995), Caudill et al. (1995), Wang (2002) and Amsler et al. (2013) and the references therein). However, these

[1] For brevity, we consider stochastic production (or or revenue or profit) frontier. For cost frontier, a simple modification by changing $-u_i$ to $+u_i$ in (1)

extensions are beyond the scope of this paper and we leave them for future exploration.

Model (1) is quite general and flexible, and it nests other stochastic frontier models as special cases. For instance, when $\rho = 0$, it reduces to Fan et al. (1996), Kumbhakar et al. (2007) models and when $m(x_i) = x_i\beta$, it becomes Glass et al. (2016) model. Finally, when $\rho = 0$ and $m(x_i) = x_i\beta$, model (1) reduces to the well known and seminal stochastic frontier model introduced by Aigner et al. (1977) and Meuseen and van den Brock (1977).

Let $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \frac{\sigma_u}{\sigma_v}$, then under our distributional assumptions of v_i and u_i, the log-likelihood function associated with equation (1) is given by:

$$\mathcal{L}(y|\rho, \lambda, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) + \log(|\boldsymbol{I}_n - \rho \boldsymbol{W}_n|) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \rho\sum_{i\neq j} w_{ij}y_j - m(x_i)\right) + \sum_{i=1}^{n}\log\left[1 - \Phi\left(\frac{\lambda\{y_i - \rho\sum_{i\neq j} w_{ij}y_j - m(x_i)\}}{\sigma}\right)\right], \tag{2}$$

where $\boldsymbol{I}_n$ is an identity matrix of dimension n, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and $\log(|\boldsymbol{I}_n - \rho\boldsymbol{W}_n|)$ is the scaled logged determinant of the Jacobian resulting from the transformation of $v_i - u_i$ to y_i. Now, since $m(\cdot)$ is unknown, equation (2) is not feasible to maximize. One approach is to use a local linear approximation for $m(\cdot)$ as in, for example, Kumbhakar et al. (2007), and maximize the local log-likelihood function associated with (2). However, there are two disadvantages of using such approach. First, since $(\rho, \lambda, \sigma^2)$ are constant parameters, the resulting local maximum likelihood will produce the estimates of these parameters that will be dependent on the localization points of x_i. Of course, one can use the profiling approach to obtain the estimates of these parameters that are constants. And second, maximizing the local log-likelihood associated with (2) can be difficult, computationally intensive and prohibitively expensive, even without imposing constraints on $m(\cdot)$.

Alternatively, one could first, consistently estimate ρ and the unknown function $m(\cdot)$ using two-stage approach in the spirit of Su (2010), and then maximizing (2) with respect to (λ, σ^2) by replacing ρ and $m(\cdot)$ by their consistent estimates. Details description of this alternative approach is given in the next section.

3 Estimation Procedure

Given the log-likelihood function (2), we propose a three-step estimation algorithm. The main idea of the three-step procedure is that, the spatial parameter ρ and the unknown frontier $m(\cdot)$ can be consistently estimated using a semiparametric profile GMM procedure in the spirit of Su (2010), in the first two steps. In the third step, the remaining parameter, namely, σ^2 and λ can be estimated by maximizing the log-likelihood function (2) by replacing ρ and $m(\cdot)$ by their consistent estimates obtained from the first two steps. Unlike the local MLE approach, by using the three-step procedure, the limiting distributions of the proposed estimators for both the parametric and nonparametric components of the model can be easily derived.

For the sake of arguments and simplicity, we first consider the estimation problem for the unconstrained case where there are no economic restrictions being imposed on the frontier $m(\cdot)$.

3.1 Unconstrained three-step estimation algorithm

Let $\epsilon_i = v_i - (u_i - E(u_i|x_i)) = v_i - u_i^*$, where under our assumption, $E(u_i|x_i) = E(u_i) = \sqrt{2/\pi}\sigma_u = \sqrt{2/\pi}\lambda\sigma/(1+\lambda^2)^{1/2}$. Then (1) can be rewritten as:

$$y_i = \rho \sum_{j \neq i} w_{ij} y_j + m^*(x_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{3}$$

where $m^*(x_i) = m(x_i) - \sqrt{2/\pi}\sigma_u$ and $E(\epsilon_i|x_i) = 0$. For estimation purpose, it is more convenient to express (3) as:

$$Y = \rho \boldsymbol{W}_n Y + \boldsymbol{m}^*(X) + \epsilon, \tag{4}$$

where $Y = (y_1, \cdots, y_n)'$ is an $(n \times 1)$ vector, $X = (x_1, \cdots, x_n)'$ is an $(n \times d)$ matrix with $x_i = (x_{i,1}, \cdots, x_{i,p})'$, $\boldsymbol{m}^*(X) = (m^*(x_1), \cdots, m^*(x_n))'$, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$ is an n-dimensional column vector.

Equation (4) resembles the semiparametric SAR model considered by Su (2010), and hence, his two-step semiparametric GMM estimation procedure can be used to estimate the unknown frontier function $m^*(\cdot)$ and the spatial coefficient ρ. To this end, let $\tilde{Y} = \boldsymbol{W}_n Y$ and $\tilde{X} = \boldsymbol{W}_n X$ where the i^{th} element of $\tilde{Y}$ and the i^{th} row of $\tilde{X}$ are denoted as $\tilde{y}_i$ and $\tilde{x}_i$, respectively. When $\rho \neq 0$, the term $\tilde{Y}$ in equation (4) is endogenously generated, and thus, we assume that there exists a $(l \times 1)$ (where $l \geq d + 1$) vector on nonstochastic instruments z_i for $\bar{x}_i = (x_i', \tilde{y}_i)'$ such that we have the following orthogonality condition:

$$E(z_i \epsilon_i) = 0. \tag{5}$$

In what follows, we assume that z_i contains a constant term. An example of such instrument is $z_i = (1, x_i', \tilde{x}_i')'$. The moment condition in (5) implies that

$$E\left[z_i \left\{y_i - \rho \tilde{y}_i - m^*(x_i)\right\}\right] = 0. \tag{6}$$

Equation (6) clearly provides moment restrictions that can be used to construct a GMM estimation procedure similar to that for parametric models. However, because $m^*(x_i)$ is unknown, the moment restrictions in (6) is not feasible to use in practice. To make (6) operational, we can approximate $m^*(x_i)$ by using local polynomial fitting approach of Fan (1992) and Fan and Gijbels (1996), albeit other approaches such as sieve can also be used in the approximation of $m^*(x_i)$ (see for example, Ai and Chen (2003)). For illustration and notational simplicity, we will focus on the local linear fitting.

Now, given a fixed point x_j, and for x_i in the neighbourhood of x_j, by taking first-order

Taylor expansion of $m^*(\cdot)$, we have:

$$m^*(x_i) \approx m^*(x_j) + \nabla m^*(x_j)'(x_i - x_j), \tag{7}$$

where $\nabla m^*(x) = \partial m^*(x)/\partial x$ is a $(d \times 1)$ vector of partial derivatives. Thus, the moment restrictions in (6) can be approximated by:

$$E\left[z_i \left\{y_i - \rho\tilde{y}_i - m^*(x_j) + \nabla m^*(x_j)'(x_i - x_j)\right\}\right] \approx 0. \tag{8}$$

Note that, due to the nature of approximation in (8), we will allow the instruments z_i to be locally dependent on the point of approximation (x_j) as well as certain parameter used in the approximation. Moreover, equation (8) contains both the global parameter ρ and the nonparametric local parameter vector $\boldsymbol{\theta}(\cdot) = (m^*(\cdot), \nabla m^*(\cdot))'$, therefore, as in Su (2010), we now propose a two-step estimation procedure to estimate these parameters.

First, by treating ρ as if it were known in (8), we can profiling out the nonparametric component and consider the estimation of $\boldsymbol{\theta}(\cdot)$ (conditional on ρ). For $r = 1, \cdots, d$, let $k((x_{i,r} - x_{j,r})/h_r)$ denotes a kernel function, and $h = (h_1, \cdots, h_d)'$ be a $(d \times 1)$ vector of bandwidths. Also, define $K_{h,i}(x_j) = \prod_{r=1}^{d} h_r^{-1} k((x_{i,r} - x_{j,r})/h_r)$ as a product kernel. Given the observed sample, it follows from (8), we have the following locally weighted orthogonality conditions:

$$Z_h(x_j)'\mathcal{K}_h(x_j)\left\{Y - \rho\tilde{Y} - Q_h(x_j)'\boldsymbol{\theta}_h(x_j)\right\} \approx 0. \tag{9}$$

where $Z_h(\cdot) = (z_{h,1}(\cdot), \cdots, z_{h,n}(\cdot))'$, and $z_{h,i}(\cdot)$ is a vector of "local instruments" which depend on the smoothing parameter h and the local approximation point x_j, $\mathcal{K}_h(.) = diag(K_{h,1}(\cdot), \cdots, K_{h,n}(\cdot))$, $Q_h(\cdot) = (q_{h,1}(\cdot), \cdots, q_{h,n}(\cdot))'$ with $q_{h,i}(\cdot) = (1, ((x_i - x_j)/h)')'$, and $\boldsymbol{\theta}_h(\cdot) = (m^*(\cdot), (h \odot \nabla m^*(\cdot))')'$ is a $((d+1) \times 1)$ vector of local parameters and $\odot$ denotes the Hadamard product. For the choice of local instruments in practice, one can simply use $z_{h,ij} = z_{h,i}(x_j)$ as:

$$z_{h,ij} = \begin{pmatrix} z_i^{(1)} \\ z_i^{(1)} \otimes ((x_i - x_j)/h) \end{pmatrix} \tag{10}$$

where $z_i^{(1)}$ is a subset of z_i and $\otimes$ denotes the Kronecker product. For example, one can set $z_i^{(1)} = 1$ or $z_i^{(1)} = z_i$ (see Su (2010) for more discussion on the choice of $z_i^{(1)}$).

Equation (9) provides the sample moment conditions that allow us to construct the following (local) GMM objective function:

$$\begin{aligned} J_\rho = & \left[Z_h(x_j)'\mathcal{K}_h(x_j)\left\{Y - \rho\tilde{Y} - Q_h(x_j)'\boldsymbol{\theta}_h(x_j)\right\}\right]' Z_h(x_j)'\mathcal{K}_h(x_j) \\ & \left\{Y - \rho\tilde{Y} - Q_h(x_j)'\boldsymbol{\theta}_h(x_j)\right\} \end{aligned} \tag{11}$$

Minimizing (11) with respect to $\boldsymbol{\theta}(\cdot)$ yields:

$$\tilde{\boldsymbol{\theta}}_{\rho,h}(x_j) = \left[\boldsymbol{A}_h(x_j)'\boldsymbol{A}_h(x_j)\right]^{-1}\boldsymbol{A}_h(x_j)'\boldsymbol{B}_{\rho,h}(x_j), \tag{12}$$

where $\boldsymbol{A}_h(x_j) = Z_h(x_j)'\mathcal{K}_h(x_j)Q_h(x_j)$, $\boldsymbol{B}_{\rho,h}(x_j) = Z_h(x_j)'\mathcal{K}_h(x_j)(Y - \rho\tilde{Y})$, and the subscript ρ that appears in (12) indicates that these quantities depend on ρ. The estimate of $m^*(x_j)$ is then given by:

$$\tilde{m}^*_{\rho,h}(x_j) = e_1'\tilde{\boldsymbol{\theta}}_{\rho,h}(x_j) = \boldsymbol{s}_h(x_j)'(Y - \rho\tilde{Y}), \tag{13}$$

where e_1 is a $((d+1)\times 1)$ vector with the first element equals to 1 and zero elsewhere, and $\boldsymbol{s}_h(x_j) = e_1'\left[\boldsymbol{A}_h(x_j)'\boldsymbol{A}_h(x_j)\right]^{-1}\boldsymbol{A}_h(x_j)'Z_h(x_j)'\mathcal{K}_h(x_j)$.

Next, given the estimate $\tilde{m}^*(.)$, we estimate the global spatial parameter ρ by parametric GMM approach. To do this, let $\boldsymbol{S}_h = ((\boldsymbol{s}_h(\boldsymbol{x}_1).\cdots,\boldsymbol{s}_h(\boldsymbol{x}_n))'$, $\bar{Y} = (\boldsymbol{I}_n - \boldsymbol{S}_h)Y$, and $Y^* = (\boldsymbol{I}_n - \boldsymbol{S}_h)\tilde{Y}$. Also, let $\boldsymbol{\Omega}_n$ be a $(l\times l)$ symmetric and positive semi-definite weighting matrix for large n. Then, the estimate of ρ can be obtained as the solution to the following minimization problem:

$$\min_\rho (\bar{Y} - \rho Y^*)' Z_h \boldsymbol{\Omega}_n Z_h'(\bar{Y} - \rho Y^*), \tag{14}$$

which yields

$$\hat{\rho} = \hat{\rho}(\boldsymbol{\Omega}_n) = \frac{Y^{*\prime} Z_h \boldsymbol{\Omega}_n Z_h^{\prime} \bar{Y}}{Y^{*\prime} Z_h \boldsymbol{\Omega}_n Z_h^{\prime} Y^*}, \tag{15}$$

where the second equality in (15) indicates that the estimate $\hat{\rho}$ depends on the choice of the weighting matrix $\boldsymbol{\Omega}_n$. As we will show later, the optimal choice of $\boldsymbol{\Omega}_n$ that minimizes the asymptotic variance of $\hat{\rho}(\boldsymbol{\Omega}_n)$ is given by $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^-$ where $\boldsymbol{\Sigma} = \sigma^2 \lim_{n\to\infty} n^{-1} Z_h^{\prime}(\boldsymbol{I}_n - \boldsymbol{S}_h)(\boldsymbol{I}_n - \boldsymbol{S}_h)^{\prime} Z_h$ and $\boldsymbol{\Sigma}^-$ is the Moore-Penrose generalized inverse of $\boldsymbol{\Sigma}$.

After we obtain estimate $\hat{\rho}$, the estimator of $\boldsymbol{\theta}_{\rho,\tilde{h}}(x_j)$ can be constructed by $\hat{\boldsymbol{\theta}}_{\tilde{h}}(x_j) = \tilde{\boldsymbol{\theta}}_{\hat{\rho},\tilde{h}}(x_j)$, and $\hat{m}^*_{\tilde{h}}(x_j) = \tilde{m}^*_{\hat{\rho},\tilde{h}}(x_j) = \boldsymbol{s}_{\tilde{h}}(x_j)^{\prime}(Y - \hat{\rho}\tilde{Y})$. Note that, we allow for different choices of the bandwidths h used in the estimation of ρ and $\boldsymbol{\theta}_{\tilde{h}}(x_j)$.

Finally, once the estimates of ρ and $m^*(x_j)$ are obtained, the remainder parameters, namely (σ^2, λ) can be estimated by maximize the log-likelihood function in (2) where we replace ρ by $\hat{\rho}$ and $m(x_j)$ by $\hat{m}^*_{\hat{\rho},\tilde{h}}(x_j) + \sqrt{2/\pi}\lambda\sigma/(1+\lambda^2)^{1/2}$, i.e.,

$$(\hat{\sigma}, \hat{\lambda}) = \underset{(\sigma,\lambda)}{\operatorname{argmax}}\ \mathcal{L}\left(\sigma, \lambda \mid \hat{\rho}, \hat{m}^*_{\hat{\rho},\tilde{h}}(x_j)\right), \tag{16}$$

and the estimate of $m(x_j)$ is given by $\hat{m}(x_j) = \hat{m}^*_{\hat{\rho},\tilde{h}}(x_j) + \sqrt{2/\pi}\hat{\lambda}_n\hat{\sigma}_n/(1+\hat{\lambda}_n^2)^{1/2}$.

Practical Implementation of the Proposed Three-Step Estimator:

To implement the proposed three-step estimator in practice, we need to choose a kernel function and the smoothing parameters h and $\tilde{h}$. For the kernel function, one can use the product of Gaussian kernels $K(x) = \prod_{r=1}^{d}(2\pi)^{-1/2}\exp(-x_r^2/2)$. For the smoothing parameters, it is difficult to determine the optimal bandwidths for h and $\tilde{h}$. Thus, following Su (2010), we suggest the following procedure to choose h and $\tilde{h}$.

First, obtain a preliminary estimate of ρ, say $\bar{\rho}$ by using a rule-of-thumb bandwidth $h_r = s_{x_r} n^{-2/7}$, where s_{x_r} is the sample standard deviation of x_r, for $r = 1, \cdots, d$. Next, given the preliminary estimate $\bar{\rho}$, conduct the the least squares cross validation (LSCV) to determine

$\tilde{h}$ by:

$$\tilde{h} = \arg\min_{\tilde{h}} n^{-1} \sum_{i=1}^{n} \left(y_i - \bar{\rho}\bar{y}_i - \bar{m}_{-i}(x_i)\right)^2 \omega(x_i), \tag{17}$$

where $\bar{y}_i$ is the i^{th} element of $\bar{Y}$, $\bar{m}_{-i}(\cdot)$ is local linear estimator of $m(\cdot)$ by leaving the observation $(x_i, \bar{y}_i)$ out in the estimation procedure and by using the bandwidth h, and $\omega(\cdot)$ is a non-negative weighting function. In our simulation below, we set $\omega(x_i) = \prod_{i=1}^{n} 1(|x_{i,r} - \bar{x}_r| \leq 2s_{x_r})$ where $1(\cdot)$ is an indicator function, and $\bar{x}_r$ is the sample average of x_r.

Now, as in the case of semiparametric estimation, undersmoothing is required for h, hence we can set $h = \tilde{h}n^{-2/7+1/(d+4)}$ to obtain an updated estimate $\hat{\rho}$ of ρ. Finally, given $\hat{\rho}$, we obtain the estimate $\hat{m}(x_j)$ of $m(x_j)$ using equation (13).

We summarize the above estimation procedure in the following three steps:

Step 1: Obtain the initial estimate $\bar{\rho}$ of ρ using (15) with the rule-of-thumb bandwidth sequence $h = (h_1, \cdots, h_d)'$. Given $\bar{\rho}$, conduct the LSCV to obtain $\tilde{h}$ as in (17). Set $h = \tilde{h}n^{-2/7+1/(d+4)}$ to obtain an updated estimate $\hat{\rho}$ of ρ.

Step 2: Given the estimate of $\hat{\rho}$, estimate the nonparametric frontier component $m^*(x_j)$ using sequence of bandwidth $\tilde{h} = (\tilde{h}_1, \cdots, \tilde{h}_d)'$ as in (13).

Step 3: Given the estimate of $\hat{\rho}$ and $\hat{m}^*(x_j)$, obtain the estimates of the remaining parameters $\gamma \equiv (\sigma^2, \lambda)'$ using (16). Using the estimates of $\hat{m}^*(x_j)$ and $\hat{\gamma} \equiv (\hat{\sigma}^2, \hat{\lambda})'$ to recover the estimate of $m(x_j)$ by $\hat{m}(x_j) = \hat{m}^*(x_j) + \sqrt{2/\pi}\hat{\lambda}_n\hat{\sigma}_n/(1 + \hat{\lambda}_n^2)^{1/2}$.

We will denote the above three-step estimator as *unconstrained semiparametric GMM* (U-SPGMM) estimator.

3.2 Constrained three-step estimation procedure

Insofar, the discussion on the estimation procedure in previous subsection does not impose any economic restrictions on $m(x_i)$. Since we are estimating a production function, theoretical properties of production function such as monotonicity and concavity, need to be incorporated in the estimation procedure[2]. This is particularly important if one interested in returns to scale (RTC) or technical change (TC). For instance, by definition, RTC is the sum of input elasticities which are to be non-negative, and it is important that these restrictions are satisfied at all data points. In this subsection, we modify the three-step estimation procedure described in the last subsection to allow for such restrictions. The modifications of the steps are as follows:

Step 1: Estimation of ρ is the same as Step 1 in the unconstrained setting, and we denote this estimate as $\hat{\rho}_c$.

Step 2: In this step, we obtain the restricted estimator of $m^*(x_j)$ by first, rewrite the unconstrained estimate $\tilde{\boldsymbol{\theta}}^*_{\rho,h}(x_j)$ in (12) as[3]:

$$\tilde{\boldsymbol{\theta}}^*(x_j) = \sum_{i=1}^{n} \mathcal{A}_{i,h}(x_j)(y_i - \hat{\rho}\bar{y}_i), \tag{18}$$

where $\mathcal{A}_{i,h}(x_j) = \left[A_{i,h}(x_j)' A_{i,h}(x_j)\right]^{-1} A_{i,h}(x_j)' z_{i,h}(x_j)' \mathcal{K}_{i,h}(x_j)$, with $A_{i,h}(\cdot)$, $z_{i,h}(\cdot)$ and $\mathcal{K}_{i,h}(\cdot)$ being the i^{th} row of $\boldsymbol{A}_h(\cdot)$, $Z_h(\cdot)$ and $\mathcal{K}_h(\cdot)$, respectively. Next, let $\boldsymbol{p} = (p_1, ..., p_n)$ and $\boldsymbol{p}_u = (\frac{1}{n}, ..., \frac{1}{n})$, then, in the spirit of Du at al. (2013) and Parmeter and Racine (2013), the constrained estimator of $m^*(x_j)$ is based on the solution $(\hat{p}_1, ..., \hat{p}_n)$ of a standard quadratic programming problem in which the objective function $D(\boldsymbol{p}) = (\boldsymbol{p}-\boldsymbol{p}_u)'(\boldsymbol{p}-\boldsymbol{p}_u)$ is minimized subject to the following relevant frontier bound, monotonicity and concavity constraints:

(i) $\sum_{i=1}^{n} p_i \mathcal{A}_{i,h}(x_j)\vec{y}_i - \vec{y}_i \geq 0$

(ii) $\sum_{i=1}^{n} p_i \mathcal{A}^{(\boldsymbol{s}_1)}_{i,h}(x_j)\vec{y}_i \geq 0$, for $\boldsymbol{s}_1 \in \boldsymbol{S}_1$,

(iii) $\sum_{i=1}^{n} p_i \mathcal{A}^{(\boldsymbol{s}_2)}_{i,h}(x_j)\vec{y}_i \leq 0$ for $\boldsymbol{s}_2 \in \boldsymbol{S}_2$,

where $\vec{y}_i = y_i - \hat{\rho}_c\bar{y}_i$, $\mathcal{A}^{(\boldsymbol{s}_r)}_{i,h}(\cdot)$ is the r^{th} order partial derivative of $\mathcal{A}_{i,h}(\cdot)$ with respect to the k-th element of x_i, for $k = 1, ..., d$, and $\boldsymbol{S}_1 = \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, \ldots, 1)\}$; whilst $\boldsymbol{S}_2 = \{(2, 0, \ldots, 0), (0, 2, \ldots, 0), \ldots, (0, \ldots, 2)\}$. Given the solution $(\hat{p}_1, ..., \hat{p}_n)$, we can obtain the restricted estimator of $m^*(x_j)$ as $\hat{m}^*_c(x_j) = e_1' \sum_{i=1}^{n} \hat{p}_i \mathcal{A}_{i,h}(x_j)\vec{y}_i$.

[2] Other theoretical axioms of producer theory (revenue, profit, cost, etc.) can also be imposed accordingly.

[3] For notational simplicity, we suppress the subscripts ρ and h.

Note that, if necessary, Steps 1 and 2 can be iterated to obtain more efficient estimators $\hat{\rho}_c$ and $\hat{m}_c^*(x_j)$ of ρ and $m^*(x_j)$, respectively.

Step 3: This step follows similarly to the unconstrained case, except now we replace $m^*(x_j)$ by $\hat{m}_c^*(x_j)$ and $\hat{m}(x_j)$ by $\hat{m}_c(x_j) = \hat{m}_c^*(x_j) + \sqrt{2/\pi}\hat{\lambda}_c\hat{\sigma}_c/(1+\hat{\lambda}_c^2)^{1/2}$, where the subscript c indicates estimates that are obtained using constrained method.

We will denote the above three-step estimator as *constrained semiparametric GMM* (C-SPGMM) estimator.

Remarks: The following remarks regarding the constraint qualifications of (ii) and (iii) are in order:

1. Imposing the monotonicity condition (ii) requires the first-order derivatives of $\mathcal{A}_{i,h}(\cdot)$, and since we are using local linear method, these estimated derivatives are readily available. However, as pointed out by Parmeter and Racine (2013), the estimated derivatives may not be exactly correspond to the *analytical* derivatives from direct calculation of $\tilde{m}^*(\boldsymbol{x}_j)$ from (18) in finite samples, for non-zero finite h. Consequently, the local linear estimator may result in a surface that in fact inconsistent with constraints. However, this is a finite sample issue and it can rectified by using the analytical derivatives instead of the estimated ones.
2. The concavity constraint in (iii) is necessary but may not be sufficient condition, unless $d = 1$. If one wants to ensure the sufficiency condition is met when $d \geq 2$, one can replace (iii) with the following conditions:

$$m^*(x_i) - m^*(x_j) \leq \frac{\partial m^*}{\partial x_1}(x_j)(x_{i,1} - x_{j,1}) + \cdots, \frac{\partial m^*}{\partial x_d}(x_j)(x_{i,d} - x_{j,d}), \ \ \forall x_i, x_j. \tag{19}$$

Afriat (1967) shows that (19) is both necessary and sufficient conditions for $m^*(\cdot)$ to be global concave, and these inequalities can be handled easily in our framework, albeit the total number of inequalities that must be imposed is $n(n-1)d$ which can be large for large n.

3.3 Estimation of marginal effects and efficiency

To obtain the marginal effects (or elasticities) and technical efficiency prediction, from (4), assuming that $\boldsymbol{D}_n = (\boldsymbol{I}_n - \rho \boldsymbol{W}_n)$ is non-singular, we can write:

$$Y = \boldsymbol{D}_n^{-1}\boldsymbol{m}(X) + \boldsymbol{D}_n^{-1}v - \boldsymbol{D}_n^{-1}u, \tag{20}$$

where $\boldsymbol{m}(X) = (m(x_1), \cdots, m(x_n))'$, $v = (v_1, \cdots, v_n)'$ and $u = (u_1, \cdots, u_n)'$. In the spirit of LeSage and Pace (2009) and Elhorst (2014), the direct, indirect and total marginal effects (ME) of y_i with respect to $x_{k,r}$ $(r = 1, \ldots, d)$ can be computed as:

$$ME_{k,r} = \frac{\partial y_i}{\partial x_{k,r}} = [\boldsymbol{D}_n^{-1}]_{ik}\frac{\partial m(x_j)}{\partial x_{k,r}}, \quad k = 1, \ldots, n, \tag{21}$$

where $x_{k,r}$ is the r^{th} frontier variable of productive unit k, and $[\boldsymbol{D}_n^{-1}]_{ik}$ is the $(i,k)^{th}$ element of $\boldsymbol{D}_n^{-1}$. The direct and indirect ME are obtained by setting $i = k$ and $i \neq k$, respectively. The total ME (TME) of the r^{th} frontier variable is defined as the partial change in y_i as response to changes in $x_{k,r}$ for all k. It is given by:

$$TME_r = \sum_{k=1}^{n} ME_{j,l} = \sum_{k=1}^{n}[\boldsymbol{D}_n^{-1}]_{ik}\frac{\partial m(x_j)}{\partial x_{k,r}}, \quad r = 1, \ldots, d, \tag{22}$$

and the total ME can be decomposed as the sum of direct and indirect ME. The estimates of these ME are obtained by replace $\boldsymbol{D}_n$ by $\hat{\boldsymbol{D}}_n = (\boldsymbol{I}_n - \hat{\rho}\boldsymbol{W}_n)$ and the partial derivatives by the estimated derivatives that come directly from $\hat{\boldsymbol{\theta}}(x_j)$. As for the interpretations of indirect marginal effects, see Glass et al. (2016).

By using similar approach as in estimating marginal effects, we can compute the direct, indirect and total inefficiency (or efficiency). The last term on the right hand side of (20) implies that $u^{TOT} = \boldsymbol{D}_n^{-1}u$, where u^{TOT} is the $(n \times 1)$ vector of total inefficiencies. It follows that the $(n \times 1)$ vector of total efficiencies is given by $\xi^{TOT} = \exp(-u^{TOT}) = \exp(-\boldsymbol{D}_n^{-1}u)$. To obtain the estimates of total efficiencies, we need the estimate of u. A popular approach to estimate u is to use Jondrow et al. (1982) method to predict u_i conditional on ε_i:

$$\hat{u}_i = \mathrm{E}(u_i|\hat{\varepsilon}_i) = \frac{\hat{\sigma}_v\hat{\sigma}_u}{\hat{\sigma}}\left(\frac{\hat{\phi}_i}{1-\hat{\Phi}_i} - \frac{\hat{\lambda}\hat{\varepsilon}_i}{\hat{\sigma}}\right), \tag{23}$$

where $\hat{\varepsilon}_i = y_i - \hat{\rho}\sum_{i\neq j} w_{ij}y_j - \hat{m}(x_i)$, $\hat{\phi}_i = \phi\left(-\frac{\hat{\lambda}\hat{\varepsilon}_i}{\hat{\sigma}}\right)$ and $\hat{\Phi}_i = \Phi\left(-\frac{\hat{\lambda}\hat{\varepsilon}_i}{\hat{\sigma}}\right)$, with $\phi(\cdot)$ being a standard normal density function, and $\hat{\rho}$, $\hat{m}(\boldsymbol{x}_i)$, $\hat{\sigma}$ and $\hat{\lambda}$ being the estimates obtained from either the U-SPGMM or the C-SPGMM method.

The estimate of k^{th} productive unit of total inefficiency $\hat{u}_k^{TOT}$ can be decomposed into direct inefficiency ($\hat{u}_k^{Dir}$) and indirect inefficiency ($\hat{u}_k^{Indir}$) as:

$$\hat{u}_k^{TOT} = \hat{u}_{k1}^{Dir} + \hat{u}_{k2}^{Indir} + \cdots + \hat{u}_{kn}^{Indir}, \quad k = 1, \ldots, n, \tag{24}$$

where we define the direct inefficiency of k^{th} productive unit as the part of the inefficiency that is resulting from reasons other than spillovers, and the indirect inefficiency as part of inefficiency that is resulting purely from spillovers of other productive units. Similarly, the estimate of total efficiency of the k^{th} productive unit, $\hat{\xi}_k^{TOT}$, can be defined as:

$$\hat{\xi}_k^{TOT} = \exp(-\hat{u}_{k1}^{Dit}) \times \exp(-\hat{u}_{k2}^{Indir}) \times \cdots + \times \exp(-\hat{u}_{kn}^{Indir}), \ \ k = 1, \ldots, n. \tag{25}$$

The estimated shares of direct and indirect inefficiencies (see for example, Kutlu (2018) and Kutlu et al. (2020)) are given by:

$$\begin{aligned} S\hat{I}E_i^{Dir} &= \frac{\hat{u}_{ii}^{Dir}}{\hat{u}_i^{TOT}} \\ S\hat{I}E_i^{Indir} &= \frac{\sum_{i\neq k}\hat{u}_{ik}^{Indir}}{\hat{u}_i^{TOT}}. \end{aligned} \tag{26}$$

The estimated shares of direct and indirect efficiencies can also be done in similar manner.

4 Limiting Theory

In this section, we will study the limiting properties of $\hat{\boldsymbol{\theta}}(x_i)$, $\hat{\rho}$ and $\hat{\gamma} = (\hat{\sigma}^2, \hat{\lambda})$ for both U-SPGMM and C-SPGMM methods. To this end, we introduce some additional notations. Let $\boldsymbol{\Psi}_n = n^{-1}Z'(\boldsymbol{I}_n - \boldsymbol{S}_h)\boldsymbol{W}_n Y$ be a $(l \times 1)$ vector, $\boldsymbol{\Psi} = \lim_{n\to\infty} n^{-1}Z'(\boldsymbol{I}_n - \boldsymbol{S}_h)\boldsymbol{W}_n Y$, $|h| = \prod_{r=1}^{d} h_r$ and $\|h\|$ denotes the Euclidean norm of h.

We make the following assumptions:

Assumption 1: (i) The spatial weighting matrix W_n has zeros diagonal elements, and the spatial parameter $\rho \in (-c_{1n}, c_{2n})$ where $0 < c_{1n}, c_{2n} < c_n < \infty$. (ii) The matrix $\boldsymbol{D}_n(\rho)$ is non-singular for all $\rho \in (c_{1n}, c_{2n})$. (iii) The row and column sum of the sequence matrices $\{\boldsymbol{W}_n\}$ and $\{\boldsymbol{D}_n(\rho)\}$ are uniformly bounded in absolute value.

Assumption 2: The error term $\{\epsilon_i = v_i - u_i^* : 1 \le i \le n, n \ge 1\}$ are independent and satisfy: $E(\epsilon_i) = 0$; $E(\epsilon_i^2) = \sigma^2 < \infty$; and $\sup_{1\le i\le n, n\ge 1} E|\epsilon_i|^{4+\delta}$ for small $\delta > 0$.

Assumption 3: (i) $\{x_i : 1 \le i \le n, n \ge 1\}$ are i.i.d with density $f_n(x)$ and $x_i \in \mathcal{X}_n \subset \mathcal{R}^d$. $f_n(x)$ is continuous and uniformly bounded on $\mathcal{X}_n$ and $\lim_{n\to\infty} f_n(x) = f(x)$ exists. (ii) There exists a continuous function $\varphi_n(x)$ such that $\lim_{n\to\infty} \varphi_n(x) = \varphi(x)$ exists, and

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \boldsymbol{z}_i^{(1)} g_n(x_i) = \lim_{n\to\infty} \int_{\mathcal{X}_n} \varphi_n(x) g_n(x) f_n(x) dx,$$

hold with probability one, for any bounded, continuous function $g_n(x)$. (iii) $m(x_i)$ is twice continuously differentiable in x, and $\sup_{x\in\mathcal{X}} m(x) \le c_m < \infty$.

Assumption 4: (i) The kernel function $k(\cdot)$ is a continuous symmetric function. There exists some small constants $\eta > 0$ and $c_k < \infty$ such that $\int |\zeta k(\zeta)|^{2+\eta} d\zeta < c_k$, $\int k^4(\zeta) d\zeta < c_k$, $\sup_\zeta k(\zeta) < c_k$, and $\sup_\zeta |\zeta| k(\zeta) < c_k$. (ii) As $n \to \infty$, $\|h\| \to 0$, $n|h| \to \infty$, and $n\|h\|^4 \to 0$. (iii) As $n \to \infty$, $\|\tilde{h}\| \to 0$, $n|\tilde{h}| \to \infty$, and $n|\tilde{h}|\|\tilde{h}\|^4 \to c_0 \in [0, \infty)$.

Assumption 5: (i) $\boldsymbol{\Omega}_n = \boldsymbol{\Omega} + o_p(1)$ where $\boldsymbol{\Omega}$ is positive semidefinite. (ii) The elements z_i of Z are uniformly bounded such that $\sup_{1\le i\le n, n\ge 1} \|z_i\| \le c_z < \infty$, and $\boldsymbol{\Psi}_n = \boldsymbol{\Psi} + o_p(1)$ with $\boldsymbol{\Psi}'\boldsymbol{\Omega}\boldsymbol{\Psi} > 0$. (iii) $\boldsymbol{\Sigma} = \sigma^2 \lim_{n\to\infty} n^{-1} Z_h'(\boldsymbol{I}_n - \boldsymbol{S}_h)(\boldsymbol{I}_n - \boldsymbol{S}_h)' Z_h$ exists and $\boldsymbol{\Psi}'\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}\boldsymbol{\Psi} > 0$.

The above assumptions are paralleled to Assumptions 1-5 of Su (2010) with some modifications since our model restricts the regressors to be continuous only, whilst Su (2010) considers both continuous and discrete regressors. Assumption 1 concerns mainly with the essential properties of spatial weights matrix. Assumption 1 (i) is a normalization rule and define the parameter space of ρ. Assumption 1 (ii) ensures that the reduced form in (20) exists, and Assumption 1 (iii) limits the spatial correlation to some degree. To simplify the computation in practice, one could use a row normalized spatial weights matrix and $\rho \in (1/r_{min}, 1)$ where r_{min} is the most negative real eigenvalue of $\boldsymbol{W}_n$. Assumption 2 restricts the error term ϵ to be i.i.d. and gives some moment conditions on ϵ. This assumption can be relaxed to allow for heteroskedasticity by allowing for the variance u to depend on a set of environmental variables. Assumption 3 is typical in nonparametric regression with continuous regressors. Assumption 4 concerns with the choice of the kennel function and the choice of the bandwidths for the first two stages estimation. They are fairly standard in nonparametric regression with continuous regressors. Assumption 5 (i) allows $\mathbf{\Omega}_n$ to be estimated from the data. Assumption 5 (ii) is standard in the spatial econometric literature, and finally Assumption 5 (iii) allow $\mathbf{\Sigma}$ to be positive semidefinite which holds if z_i contains x_i.

4.1 Limiting properties of U-SPGMM estimator

We first study the limiting distribution of the U-SPGMM estimator of $\hat{\rho}$, $\hat{\boldsymbol{\theta}}(x)$, and $(\hat{\sigma}^2, \hat{\lambda})$. Let $\kappa_{ij} = \int a^i k^j(a) da$ for $i, j = 0, 1, 2$,

$$\boldsymbol{\mathcal{B}}(x) = \begin{pmatrix} \varphi(x) & \mathbf{0}_{l_1 \times d} \\ \mathbf{0}_{l_1 d \times 1} & \kappa_{21}\varphi(x) \otimes \boldsymbol{I}_d \end{pmatrix}$$

$$\mathbf{\Gamma}_n(x) = n^{-1}\sigma^2 \prod_{r=1}^{d} \tilde{h}_r \left[\mathcal{K}_{\tilde{h}}(x) Z_{\tilde{h}}(x)\right]' \left[\mathcal{K}_{\tilde{h}}(x) Z_{\tilde{h}}(x)\right]$$

where l_1 is the dimension of $z_i^{(1)}$. Assuming that $\lim_{n\to\infty}\mathbf{\Gamma}_n = \mathbf{\Gamma}$ exists and x is an interior point of $\mathcal{X}$, the we have the following result.

Theorem 4.1. *Under Assumptions 1-5, we have,*

(i) $\sqrt{n}(\hat{\rho}-\rho) \xrightarrow{d} N\left(0, (\mathbf{\Psi}'\mathbf{\Omega}\mathbf{\Psi})^{-2}\mathbf{\Psi}'\mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}\mathbf{\Psi}\right)$,

(ii) $\sqrt{n|\tilde{h}|}\left(\hat{\boldsymbol{\theta}}_{\tilde{h}}(x) - \boldsymbol{\theta}_{\tilde{h}}(x) - Bias(\tilde{h}, x)\right) \xrightarrow{d} N\left(0, f^{-2}(x)\boldsymbol{\mathcal{B}}^*(x)\mathbf{\Gamma}(x)\boldsymbol{\mathcal{B}}^*(x)'\right)$,

(iii) $\sqrt{n}(\hat{\gamma}-\gamma) \xrightarrow{d} N(0, \boldsymbol{\mathcal{I}}^{-1})$,

where $\boldsymbol{\theta}_{\tilde{h}}(x) = \left(m^*(x), (\tilde{h}\odot\nabla m^*(x))'\right)'$, $\boldsymbol{\mathcal{B}}^*(x) = \left(\boldsymbol{\mathcal{B}}(x)'\boldsymbol{\mathcal{B}}(x)\right)^{-1}\boldsymbol{\mathcal{B}}(x)'$, *and*

$$Bias(\tilde{h}, x) = \boldsymbol{\mathcal{B}}^*(x)\begin{pmatrix} \frac{1}{2}\kappa_{21}\varphi(x)\sum_{r=1}^{d}\tilde{h}_r^2 m_{rr}^*(x) \\ \mathbf{0}_{l_1 d\times 1} \end{pmatrix}$$

with $m_{rr}^*(x) = \partial^2 m^*(x)/\partial x_r^2$, *for* $r = 1, \cdots, d$, *and* $\boldsymbol{\mathcal{I}} = E\left[\partial^2\mathcal{L}(\gamma)/\partial\gamma\partial\gamma'\right]$.

Proof. The proofs of parts (i) and (ii) in Theorem 4.1 follow closely to the proofs of Theorems 3.1 and 3.2 of Su (2010) with a few minor modifications for i.i.d. errors and no discrete regressors case. Therefore we omit it here. The proof of part (iii) follows standard arguments of the maximum likelihood theory. □

Note that, the result in Theorem 4.1 (i) implies that the optimal choice for the weighting matrix $\mathbf{\Omega}_n$ that minimizes the asymptotic variance of $\hat{\rho}$ is $\mathbf{\Omega}_n = \mathbf{\Sigma}^-$. With this optimal choice, it is easy to verify that the asymptotic variance of $\hat{\rho}(\mathbf{\Sigma}^-)$ is given by $(\mathbf{\Psi}'\mathbf{\Sigma}^-\mathbf{\Psi})^{-1}$. Furthermore, the result in Theorem 4.1 (ii) implies that

$$\sqrt{n|\tilde{h}|}\left(\hat{m}_{\tilde{h}}^*(x) - m_{\tilde{h}}^*(x) - e_1'Bias(x)\right) \xrightarrow{d} N\left(0, f^{-2}(x)e_1'\boldsymbol{\mathcal{B}}^*(x)\mathbf{\Gamma}(x)\boldsymbol{\mathcal{B}}^*(x)'e_1\right).$$

For inference purposes, we need the estimates of the asymptotic variances of $\hat{\rho}$ and $\hat{\gamma}$. Let $\hat{\epsilon} = Y - \hat{\rho}\boldsymbol{W}_n Y - \hat{\boldsymbol{m}}(X)$, $\hat{\sigma}^2 = n^{-1}\hat{\epsilon}'\hat{\epsilon}$, and $\hat{\mathbf{\Sigma}} = n^{-1}\hat{\sigma}^2 Z_h'(\boldsymbol{I}_n - \boldsymbol{S}_h)(\boldsymbol{I}_n - \boldsymbol{S}_h)'Z_h$. Then, under the optimal choice of the weighting matrix $\mathbf{\Omega}_n$, the estimated asymptotic variance of $\hat{\rho}$ can be constructed as $V\hat{a}r(\hat{\rho}) = (\mathbf{\Psi}_n'\hat{\mathbf{\Sigma}}^{-}\mathbf{\Psi})^{-1}$. For the estimated asymptotic variance of $\hat{\gamma}$, it is given by $\hat{\mathcal{I}} = n^{-1}\partial\mathcal{L}^2(\hat{\gamma})/\partial\gamma\partial\gamma'$. In some instances, we may wish to make inference on the nonparametric component as well. For this purpose, we need to estimate the asymptotic variance matrix given in Theorem 4.1 (ii). Let $\hat{\Gamma} = n^{-1}\hat{\sigma}^2\prod_{r=1}^{d}\tilde{h}_r Z_{\tilde{h}}'(x)\mathcal{K}_{\tilde{h}}(x)\left[\mathcal{K}_{\tilde{h}}(x)Z_{\tilde{h}}(x)\right]$, $\hat{f}(x) = n^{-1}\sum_{i=1}^{n}\mathcal{K}_{\tilde{h},i}(x)$, and $\hat{\varphi}(x) = n^{-1}\sum_{i=1}^{n}z_i^{(1)}\mathcal{K}_{\tilde{h},i}(x)/\hat{f}(x)$. Define $\hat{\mathcal{B}}(x)$ as $\mathcal{B}(x)$ with $\hat{\varphi}(x)$ replacing $\varphi(x)$, and $\hat{\boldsymbol{\mathcal{B}}}^*(x) = \left(\hat{\boldsymbol{\mathcal{B}}}(x)'\hat{\boldsymbol{\mathcal{B}}}(x)\right)^{-1}\hat{\boldsymbol{\mathcal{B}}}(x)'$. Then, the estimated asymptotic variance matrix of $\hat{\boldsymbol{\theta}}_{\tilde{h}}(x)$ is given by $\left(\hat{\boldsymbol{\mathcal{B}}}(x)'\hat{\boldsymbol{\mathcal{B}}}(x)\right)^{-1}\hat{\boldsymbol{\mathcal{B}}}(x)'/\hat{f}^2(x)$.

4.2 Limiting properties of C-SPGMM estimator

We now derive the limiting distribution of the constrained estimator for $m_c^*(x)$. The limiting distributions of $\hat{\rho}_c$ and $\hat{\gamma}_c$ are unaffected by the constrained estimation of $m^*(x)$ and hence they are the same as in Theorem 4.1. We need the following additional assumption.

Assumption 6: (i) The kernel function $k(v)$ satisfies $|a^j k(a) - b^j k(b)| \leq M|a-b|$ for any $a, b \in \Re$ and $0 \leq j \leq 3$. (ii) There exists $\{\hat{p}_i\}_{1\leq i\leq n}$ such that $\max_{1\leq i\leq n}|\hat{p}_i| = O_p(n^{-1})$ holds for all $x \in \mathcal{X}_n$.

Assumption 6 is needed to ensure the existence of $\hat{p}_i$ and its convergence rate. Assumption 6 (ii) can be verified for a class of nonparametric regression models including the smooth coefficient models, that are based on the local polynomial fitting method. For more discussion on the existence and convergence rate of $\hat{p}_i$, see for example, Racine et al. (2009) and Malikov and Sun (2017).

Theorem 4.2. *Under Assumptions 1-6, we have:*

$$\sqrt{n|\tilde{h}|}\left(\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \boldsymbol{\theta}_{\tilde{h}}(x) - Bias_1(\tilde{h}, x)\right) \xrightarrow{d} N(0, f^{-2}(x)\boldsymbol{\mathcal{B}}^*(x)\mathbf{\Gamma}(x)\boldsymbol{\mathcal{B}}^*(x)'),$$

where $Bias_1(\tilde{h}, x) = Bias(\tilde{h}, x) + (n|\tilde{h}|)^{-1/2}\varphi(x)$, *and* $Bias(\tilde{h}, x)$, $\boldsymbol{\mathcal{B}}^*(x)$ *and* $\mathbf{\Gamma}(x)$ *are defined in Theorem 4.1.*

Proof. (*Sketch of Proof*)

First, we write: $\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \boldsymbol{\theta}_{\tilde{h}} = \left(\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \hat{\boldsymbol{\theta}}_{\tilde{h}}(x)\right) + \left(\hat{\boldsymbol{\theta}}_{\tilde{h}}(x) - \boldsymbol{\theta}_{\tilde{h}}(x)\right)$. Next, we study the distant between $\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x)$ and $\hat{\boldsymbol{\theta}}_{\tilde{h}}(x)$. Recall that $\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) = \sum_{i=1}^{n} \hat{p}_i \mathcal{A}_{i,h}(x_j)\vec{y}_i$, and $\hat{\boldsymbol{\theta}}_{\tilde{h}}(x) = \sum_{i=1}^{n} \mathcal{A}_{i,h}(x_j)\vec{y}_i$, then, by Hölder's inequality, we have:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \hat{\boldsymbol{\theta}}_{\tilde{h}}(x) = \sum_{i=1}^{n}(n\hat{p}_i - 1)\mathcal{A}_{i,h}(x_j)\vec{y}_i &\leq \left[\sum_{i=1}^{n}(n\hat{p}_i - 1)^2\right]^{1/2} \left[\sum_{i=1}^{n}\mathcal{A}_{i,h}^2(x_j)\vec{y}_i^2\right]^{1/2} \\ &= T_{n1} + T_{n2}, \end{aligned} \tag{27}$$

where the definitions of T_{n1} and T_{n2} are apparent. By assumption 6, we have $T_{n1} = O_p(1)$, and by following the proof of Theorem 3.2 in Su (2010), it can shown that $T_{n2} = O_p\left((n|\tilde{h}|)^{-1/2}\right)$. Thus, it follows that $\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \hat{\boldsymbol{\theta}}_{\tilde{h}}(x) = O_p(1)O_p\left((n|\tilde{h}|)^{-1/2}\right) = O_p\left((n|\tilde{h}|)^{-1/2}\right)$ or $\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) = \hat{\boldsymbol{\theta}}_{\tilde{h}}(x) + O_p(1)O_p\left(n|\tilde{h}|)^{-1/2}\right)$. Consequently, we have:

$$\sqrt{n|\tilde{h}|}\left(\hat{\boldsymbol{\theta}}_{\tilde{h},c}(x) - \boldsymbol{\theta}_{\tilde{h}}(x)\right) = \sqrt{n|\tilde{h}|}\left(\hat{\boldsymbol{\theta}}_{\tilde{h}}(x) - \boldsymbol{\theta}_{\tilde{h}}(x) + O_p\left((n|\tilde{h}|)^{-1/2}\right)\right).$$

The result then follows from the proof of Theorem 3.2 in Su (2010). □

Comparing the result in Theorem 4.2 with $\hat{\boldsymbol{\theta}}_{\tilde{h}}(x)$ in Theorem 4.1 shows that the limiting distribution of $\hat{\boldsymbol{\theta}}_{\tilde{h},c}$ has an additional vanishing asymptotic bias term of order $O_p\left((n|\tilde{h}|)^{-1/2}\right)$ resulting from imposing the valid constraints. However, the two estimators has the same asymptotic variance. The result in Theorem 4.2 also implies that

$$\sqrt{n|\tilde{h}|}\left(\hat{m}^*_{\tilde{h}.c}(x) - m^*_{\tilde{h}}(x) - e_1^{'}Bias_1(x)\right) \xrightarrow{d} N(0, f^{-2}(x)e_1^{'}\boldsymbol{\mathcal{B}}^*(x)\boldsymbol{\Gamma}(x)\boldsymbol{\mathcal{B}}^*(x)^{'}e_1).$$

5 Monte Carlo Simulations

To examine the finite sample performance of the proposed estimators, we conduct a small set of Monte Carlo experiments. We generate the data from the following data generating processes (DGPs):

$$y_i = \rho \sum_{j \neq i} w_{ij} y_j + m(x_{i1}, x_{i2}) + v_i - u_i, \quad i = 1, \ldots n \tag{28}$$

where the function $m(.)$ is generated according to the following production function specifications:

DGP 1: $m(x_{i1}, x_{i2}) = 1 + x_{i1} + x_{i2}$

DGP 2: $m(x_{i1}, x_{i2}) = [0.6x_{i1}^2 + 0.4x_{i2}^2]^{1/2}$

where DGP 1 corresponds to the popular Cobb-Douglas specification, whilst DGP 2 corresponds to the constant elasticity substitution (CES) specification. It is easy to verify that both specifications satisfy the monotonicity and concavity conditions. The input variables x_j are generated as i.i.d. $\chi^2_{(1)}$, for $j = 1, 2$. The two-sided error and inefficiency are generated according to $v_i \sim N(0, \sigma_v^2)$ and $u_i \sim N^+(0, \sigma_u^2)$, respectively, and all are mutually independent. We generate the spatial weights w_{ij} using scalar-normalized exponential distance as follows.

$$w_{ij} = \begin{cases} \frac{\tilde{w}_{ij}}{\tilde{\lambda}_{\max}}, & i \neq j \\ 0, & i = j, \end{cases}$$

where $\tilde{w}_{ij} = exp(-d_{ij})$ for $i \neq j$ and 0 for $i = j$, with d_{ij} are the centroid distant between each pair of spatial units i and j, and $\tilde{\lambda}_{\max} > 0$ is the largest eigenvalue of $\tilde{W} = [\tilde{w}_{ij}]$. The distance d_{ij} is generated as $d_{ij} = |d_i - d_j|$ where d_i and d_j are drawn independently from a uniform distribution.

We set the values of $\rho = \{0, 0.4, 0.7\}$ and we use the same values of σ^2 and λ employed by Aigner et al. (1997): $(\sigma^2, \lambda) = (1.88, 1.66), (1.63, 1.24)$ and $(1.35, 0.83)$ which corresponds to $(\sigma_v^2, \sigma_u^2) = (0.500, 1.379), (0.339, 0.901)$ and $(0.294, 0.536)$, respectively[4]. For each Monte Carlo experiment, we consider samples of size $n = 200, 400$ and 800 and the number of Monte Carlo replications are 1000 for the case $n = 200, 400$, and 500 for $n = 800$.

[4]These values also used by Fan et al. (1996) and Parmeter and Racine (2013).

For both unconstrained and constrained estimators, the choice of kernel and bandwidths are discussed near the end of subsection 3.1. For the constrained estimator, our monotonicity and concavity constraints are imposed on a grid of 100 equally spaced points to reduce computational burden. For the estimates of the parametric component in the model, we report the bias and mean squares error (MSE), whilst for nonparametric component in the model, we report only the MSE. Note that, for the constrained estimator, the MSE is calculated as the average squared difference between the estimator and the true frontier on the same set of grid points use to impose the constraints.

To demonstrate the robustness of our proposed estimators against misspecification of the production frontier, we also include the parametric SAR stochastic frontier model of Glass et al. (2016).

6 Concluding Remarks