

Constrained Nonparametric Estimation of Input Distance Function

Kai Sun

the date of receipt and acceptance should be inserted later

Abstract This paper proposes a constrained nonparametric method of estimating an input distance function. A regression function is estimated via kernel methods without functional form assumptions. To guarantee that the estimated input distance function satisfies its properties, monotonicity constraints are imposed on the regression surface via the constraint weighted bootstrapping method borrowed from statistics literature. The first, second, and cross partial analytical derivatives of the estimated input distance function are derived, and thus the elasticities measuring input substitutability can be computed from them. The method is then applied to a cross-section of 3,249 Norwegian timber producers.

Keywords Nonparametric estimation · Input distance function · Constraints · Elasticities

JEL Codes C14 · D24

1 Introduction

One of the main objectives of productivity analysis is to estimate a representation of technology using econometrics or data envelopment analysis (DEA), among others. It is well-known that DEA assumes no functional form of a frontier, and linear programming allows one to impose linear constraints on each observation. However, when it comes to econometric methods, a functional form is usually required, and information (e.g., marginal cost/product) can be calculated after parameters are estimated. Researchers should then check if the estimated information follows theoretical properties of the technology. While using a large micro-level data set, however, it may well be the case that such properties are not satisfied for a subset of observations. In order to address this issue, Terrell (1996), Rambaldi and Doran (1997), Ryan and Wales (2000), and Henningsen and Henning (2009), among others, proposed methods of imposing various constraints on parametric functions. However, it is still a problem to impose these constraints on a nonparametric function.

In this paper, we provide solutions to the two above-mentioned problems of econometric methods, i.e., functional forms that are not flexible and theoretical constraints that are difficult to impose. To address the first problem, we propose estimating a technology without a parametric functional form via nonparametric kernel econometric methods. Thus, the functional form assumption is relaxed and the technology is estimated in a fully flexible manner. The price one has to pay for such flexibility, however, is many observations may violate properties of the technology. Therefore, it is desirable that constraints are imposed on the nonparametric regression function such that the properties are satisfied for each individual observation. To do this, we apply the new approach of constraint weighted bootstrapping (CWB) first introduced by Hall and Huang (2001) and further studied by Du et al (2013) and Parmeter et al (2013).

To explore the kernel and the CWB methods, we first need to specify a technology as our main focus. In this paper, we choose an input distance function (IDF) because although the functional form of an IDF is generally unknown (Färe et al, 1994), many past and more recent studies have sought to estimate parametric distance functions specifying a translog functional form and using ordinary least squares (OLS) to estimate the unknown parameters (Lovell et al, 1994; Grosskopf et al, 1997; Ray, 2003; Cuesta et al, 2009, among others). Färe et al (1985, 1994) and Coelli and Perelman (1999), among others, used DEA to estimate the distance function without specifying any functional form. This paper fills the gap by estimating the IDF using nonparametric econometric methods without any functional form assumption.¹ Furthermore, monotonicity constraints based on the properties of the IDF are imposed via CWB.² Our methodology extends to other representations of technology in a straightforward manner. As a by-product of the econometric estimation, the first and second order analytical derivatives of the nonparametric IDF are derived, and thus the elasticities measuring input substitutability/complementarity can be calculated from them.

As an empirical example, we apply the proposed methodology to a Norwegian forestry data set from Lien et al (2007) compiled by Statistics Norway. Both the unconstrained and constrained nonparametric IDFs as well as the implied

¹ Another primal representation of technology is a production function. A nonparametric production function via kernel regression has been studied by Henderson (2009), Du et al (2013), among others. Furthermore, the estimation of the production function may require fewer regularities as it is arguable that the marginal product of labor can be negative because of labor hoarding or regulation (Heshmati et al, 2013).

² An IDF is dual to a cost function, and therefore, it must satisfy conditions similar to those of the cost function. However, estimation of the IDF does not require input price data.

elasticities are estimated and results are compared. We find that without imposing constraints, 18.25% observations violate one of the theoretical properties of the IDF. The Kolmogorov-Smirnov test shows that the gradients and the elasticities calculated from the constrained IDF significantly differs from those calculated from the unconstrained counterpart. Finally, we reported density plots of the estimated Antonelli, Morishima, and Symmetric elasticities of complementarity.

The rest of the paper is organized as follows. Section 2 describes the methodology of the constrained nonparametric econometric estimation methodology. Section 3 applies the methodology to a real data set. Section 4 discusses limitations and possible extensions of the current method, and section 5 concludes the paper.

2 Methodology

2.1 Nonparametric Estimation of a Distance Function via Kernel Methods

The distance function representation of a production technology, proposed by Shephard (1953, 1970) does not require any aggregation, prices, or behavioral assumptions. Following Färe and Primont (1995), we first define the production technology of the firm using the input set, $L(Y)$, which represents the set of all K inputs, $X \in \mathbb{R}_+^K$, which can produce the vector of Q outputs, $Y \in \mathbb{R}_+^Q$. That is:

$$L(Y) = \{X \in \mathbb{R}_+^K : X \text{ can produce } Y\}. \quad (1)$$

We assume that the technology satisfies the standard axiom of strong disposability. The IDF is then defined on the input set, $L(Y)$, as:

$$D(X, Y) = \max\{\rho : (X/\rho) \in L(Y)\}, \quad (2)$$

where ρ is the scalar distance by which the input vector can be deflated. $D(X, Y)$ satisfies the following properties: (1) it is non-increasing in each output level; (2) it is non-decreasing in each input level; (3) it is homogeneous of degree 1 in X .³ It is based on an input-saving approach and gives the maximum amount by which an input vector can be radially contracted while still being able to produce the same output vector. The IDF, $D(X, Y)$, will take a value which is greater than or equal to one if the input vector, X , is an element of the feasible input set, $L(Y)$. That is, $D(X, Y) \geq 1$ if $X \in L(Y)$. The distance function will take a value of unity if X is located on the inner boundary of the input set.

To empirically estimate the distance function, we first define:

$$D \equiv A \cdot D(X, Y), \quad (3)$$

where A is the productivity parameter, and X and Y are the input and output vectors, respectively. Using property (3) of the IDF, viz., homogenous of degree 1 in X , we can write (3) as:

$$D/X_1 = A \cdot D(\tilde{X}, Y), \quad (4)$$

³ An input distance function is concave in inputs if the input requirement set, $L(Y)$, is convex (Kumbhakar and Lovell 2000, p.30). However, in this paper, we do not assume that $L(Y)$ is convex, since “a convexity property is occasionally added to the list of properties satisfied by the input sets $L(Y)$.” (Kumbhakar and Lovell 2000, p.22)

where X_1 is the numeraire input, and \tilde{X} is a vector of input ratios, with elements $\tilde{X}_k = X_k/X_1, \forall k = 2, \dots, K$. Taking the natural logarithm for both sides gives:

$$\ln D - \ln X_1 = \ln A + \ln D(\tilde{X}, Y). \quad (5)$$

Letting $D = 1$ would give:⁴

$$\begin{aligned} -\ln X_1 &= \ln D(\tilde{X}, Y) + \ln A \\ &= \ln D(\exp(\ln \tilde{X}), \exp(\ln Y)) + \ln A \\ &\equiv m(\ln \tilde{X}, \ln Y) + v, \end{aligned} \quad (6)$$

where $v = \ln A$ is the noise term, interpreted as the natural logarithm of the productivity parameter.⁵ Using general notation, (6) can be written as:

$$\mathcal{Y} = m(z) + v, \quad (7)$$

where $\mathcal{Y} = -\ln X_1$, $m(\cdot)$ is the unknown smooth distance function, z is the vector of continuous variables (i.e., $\ln \tilde{X}_k, \forall k = 2, \dots, K$; $\ln Y_q, \forall q = 1, \dots, Q$), and v is the random error uncorrelated with any element of z .

To estimate the unknown function, one can use the local-constant least-squares estimator of $m(z)$ (see Li and Racine (2006) for more details), given by:

$$\hat{m}(z) = \frac{\sum_{i=1}^n K\left(\frac{z_i - z}{h}\right) \mathcal{Y}_i}{\sum_{i=1}^n K\left(\frac{z_i - z}{h}\right)}, \quad (8)$$

where $K(\cdot)$ is a (scalar) Gaussian product kernel weighting function for the continuous variables (see Appendix B for an explicit expression); n denotes the sample size; h is a vector of bandwidth, with each element for a particular variable in the z vector.

Estimation of the bandwidths, h , is typically the most salient factor when performing nonparametric estimation. Although many selection methods exist, we utilize the data-driven least-squares cross validation (LSCV) method. Specifically, the bandwidths are chosen to minimize

$$n^{-1} \sum_{i=1}^n [\mathcal{Y}_i - \hat{m}_{-i}(z_i)]^2, \quad (9)$$

where $\hat{m}_{-i}(\cdot) = \frac{\sum_{j \neq i}^n K\left(\frac{z_j - z_i}{h}\right) \mathcal{Y}_j}{\sum_{j \neq i}^n K\left(\frac{z_j - z_i}{h}\right)}$ is the leave-one-out local-constant kernel estimator of $m(\cdot)$. We use the `npregbw` function from the `np` package (Hayfield and Racine, 2008) in R (R Development Core Team, 2011) to estimate the bandwidth vector. This bandwidth vector is then plugged into (8) to estimate the IDF.

To calculate the derivatives of the distance function with respect to each input and output, (6) can be re-written as:

$$\ln D = \ln X_1 + m(\ln \tilde{X}_2, \dots, \ln \tilde{X}_K, \ln Y_1, \dots, \ln Y_Q) + v. \quad (10)$$

⁴ Alternatively, the estimating equation can be derived by defining $D(X, Y) \equiv 1/A$, and then imposing the homogeneity restriction and taking the natural logarithm. We would like to thank an anonymous referee for the suggestion.

⁵ In fact, the translog specification of $m(\cdot)$ in (6) can be derived from a second-order Taylor expansion of $m(\cdot)$, whose gradients are viewed as unknown parameters (Berndt and Christensen, 1973). But we do not have to rely on the Taylor approximation, since we can estimate (6) using kernel-based nonparametric methods without specifying any parametric functional form.

For example, the first partial derivatives of interest, that should be investigated to guarantee the monotonicity properties of the IDF, are:

$$\frac{\partial \ln D}{\partial \ln X_1} = 1 - \sum_{k=2}^K \frac{\partial m}{\partial \ln \tilde{X}_k}, \quad (11)$$

$$\frac{\partial \ln D}{\partial \ln X_k} = \frac{\partial \ln D}{\partial \ln \tilde{X}_k} = \frac{\partial m}{\partial \ln \tilde{X}_k}, \quad \forall k = 2, \dots, K,^6 \quad (12)$$

and

$$\frac{\partial \ln D}{\partial \ln Y_q} = \frac{\partial m}{\partial \ln Y_q}, \quad \forall q = 1, \dots, Q, \quad (13)$$

where the first partial derivative of $m(\cdot)$ with respect to a particular argument, say, $z_l \in \{\ln \tilde{X}_2, \dots, \ln \tilde{X}_K, \ln Y_1, \dots, \ln Y_Q\}$, is

$$\frac{\partial m}{\partial z_l} = \sum_{i=1}^n \frac{\left(\frac{z_{li} - z_l}{h_l^2}\right) K(\cdot) \sum_i K(\cdot) - K(\cdot) \sum_i \left[\left(\frac{z_{li} - z_l}{h_l^2}\right) K(\cdot)\right]}{\left(\sum_i K(\cdot)\right)^2} \cdot (-\ln X_{1i}). \quad (14)$$

See Appendix B for detailed derivation of the first and the second partial derivatives of $m(\cdot)$ with respect to z_l , and the cross partial derivatives with respect to z_l and z_k , $\forall l \neq k$.

2.2 Imposition of Regularity Constraints

Recall that the IDF has the following theoretical properties:

$$\frac{\partial \ln D}{\partial \ln X_k} \geq 0, \quad \forall k = 1, \dots, K \quad (15)$$

and

$$\frac{\partial \ln D}{\partial \ln Y_q} \leq 0, \quad \forall q = 1, \dots, Q.^7 \quad (16)$$

However, when it comes to empirical estimation, it is very likely for one to obtain violations of these properties for some individual observation. Most empirical researchers check these regularity conditions at the mean of the data instead of every data point, and report results evaluated at the mean. This practice defeats the purpose of using micro data. Results may not be of much use for policy analysis if the theoretical restrictions are violated for many individual producers. Instead of ignoring results that violate rationality, we use a new statistical method that imposes these economic constraints. We then calculate the gradients of the IDF and the elasticities based on that all these joint constraints are satisfied.

In order to impose such observation-specific inequality constraints, we follow the constraint weighted bootstrapping (CWB) method first proposed by Hall and Huang (2001) and further studied by Du et al (2013) and Parmeter et al (2013), whose idea is to transform the response variable by assigning observation-specific weights such that certain constraints in the model are satisfied. To illustrate this methodology, let $\{\mathcal{Y}_i, z_i\}_{i=1}^n$ denote sample pairs of response and explanatory variables, where \mathcal{Y}_i is a scalar,⁸ z_i is of dimension $(K + Q - 1)$, and n denotes the sample size. The goal is to estimate the conditional mean model $\mathcal{Y} = m(z) + v$, subject to constraints on the first order gradient of the l -th element

⁶ By chain-rule, $\partial \ln D / \partial \ln X_k = \partial \ln D / \partial \ln \tilde{X}_k$, $\forall k = 2, \dots, K$.

⁷ Since D , X_k , and Y_q are all non-negative, $\frac{\partial \ln D}{\partial \ln X_k} \geq 0$ implies $\frac{\partial D}{\partial X_k} \geq 0$, and $\frac{\partial \ln D}{\partial \ln Y_q} \leq 0$ implies $\frac{\partial D}{\partial Y_q} \leq 0$.

⁸ $\mathcal{Y}_i = -\ln X_{1i}$ in this paper.

in z , $m_l(z) = \partial m(z) / \partial z_l$, where z_l is the l -th element of the vector z , or on a linear combination of any of the first order gradients.

We can express the local-constant estimator as:

$$\hat{m}(z) = n \cdot \sum_{i=1}^n A_i(z) n^{-1} \mathcal{Y}_i, \quad (17)$$

where $A_i(z) = \frac{K(\frac{z_i - z}{h})}{\sum_{i=1}^n K(\frac{z_i - z}{h})}$, $K(\cdot)$ and h are defined the same as in (8).⁹ The first order gradient of the local-constant estimator, $\hat{m}_l(z)$, can be expressed as:

$$\hat{m}_l(z) = n \cdot \sum_{i=1}^n A_{i,l}(z) n^{-1} \mathcal{Y}_i, \quad (18)$$

where $A_{i,l}(z) = \frac{\partial A_i(z)}{\partial z_l}$ (see Appendix B for an explicit expression of the derivative of $A_i(z)$ with respect to z_l). Therefore, a particular linear combination of these first order gradients is:

$$1 - \sum_{l=1}^{K-1} \hat{m}_l(z) = 1 - \sum_{l=1}^{K-1} \left(n \cdot \sum_{i=1}^n A_{i,l}(z) n^{-1} \mathcal{Y}_i \right), \quad (19)$$

which is used to impose constraints in the form of (11).

To impose the monotonicity constraints, re-write (18) and (19) as:

$$\hat{m}_l(z|p) = n \cdot \sum_{i=1}^n A_{i,l}(z) p_i \mathcal{Y}_i, \quad \forall l = 1, \dots, K + Q - 1, \text{ and} \quad (20)$$

$$1 - \sum_{l=1}^{K-1} \hat{m}_l(z|p) = 1 - \sum_{l=1}^{K-1} \left(n \cdot \sum_{i=1}^n A_{i,l}(z) p_i \mathcal{Y}_i \right), \quad (21)$$

where p_i is the weight for the i th observation of the response \mathcal{Y} , and $\sum_{i=1}^n p_i = 1$.¹⁰ If $p_i = 1/n$ (i.e., uniform weights), then the constrained estimator will reduce to the unconstrained estimator.

The goal is to transform the response as little as possible through the weights such that the constraints are satisfied. The following is the weight selection criterion proposed by Du et al (2013) and Parmeter et al (2013):

$$\begin{aligned} p^* &= \operatorname{argmin} \mathcal{D}(p) = (p_u - p)'(p_u - p) \\ \text{st. } \mathbf{l}(z) &\leq \hat{m}_l(z | p) \leq \mathbf{u}(z), \quad \forall l = 1, \dots, K + Q - 1, \text{ and} \\ \mathbf{l}(z) &\leq 1 - \sum_{l=1}^{K-1} \hat{m}_l(z | p) \leq \mathbf{u}(z), \end{aligned} \quad (22)$$

⁹ The CWB method is also applicable to the local-linear kernel estimator. This is because (17) becomes the local-linear estimator (Li and Racine, 2004) if we write

$$A_i(z) = \left\{ \sum_{i=1}^n K \left(\frac{z_i - z}{h} \right) \begin{bmatrix} 1 & z_i - z \\ z_i - z & (z_i - z)(z_i - z)' \end{bmatrix} \right\}^{-1} K \left(\frac{z_i - z}{h} \right) \begin{bmatrix} 1 \\ z_i - z \end{bmatrix}$$

Note that the $A_i(z)$ in the local-linear case is a 2×1 block vector. The first element of it is the conditional mean, and the second element gives the gradient vector.

¹⁰ Du et al (2013) showed that p_i can be either positive or negative for the purpose of imposing generalized constraints. In contrast, Hall and Huang (2001) used the power divergence metric which restricts $0 \leq p_i \leq 1$.

where p^* is a vector of optimal weight for each response observation, $\mathcal{D}(p)$ is an L_2 metric, p_u is a vector of uniform weights (i.e., $1/n$), which can also be viewed as an initial search condition, $\mathbf{l}(z)$ and $\mathbf{u}(z)$ represent observation-specific lower and upper bounds for $\hat{m}_l(z | p)$, respectively. If $\mathbf{l}(z) = 0$ and $\mathbf{u}(z) = +\infty$, then we can impose monotonically increasing constraints; if $\mathbf{u}(z) = 0$ and $\mathbf{l}(z) = -\infty$, then we can impose monotonically decreasing constraints. The optimization problem (22) is a standard quadratic programming problem that can be numerically solved using the *quadprog* package (Berwin and Weingessel, 2011) in *R*. The constrained estimator $\hat{m}(z | p^*) = n \sum_{i=1}^n A_i(z) p_i^* \mathcal{Y}_i$ can then be calculated using the optimal weight for each observation, p_i^* . Following Du et al (2013) and Parmeter et al (2013), the same bandwidth vector estimated from the unconstrained IDF are used to estimate the constrained IDF, as the same $A_i(z)$ appears in both the unconstrained and the constrained estimator. The codes for estimating the constrained model are available from the author upon request.

2.3 Elasticities from the Distance Function

After the first, second, and cross partial derivatives of the IDF satisfying theoretical restrictions are estimated, three types of elasticities measuring input substitutability/complementarity can be computed without any additional information (Stern, 2011), viz., the Antonelli elasticity of complementarity (AEC), the Morishima elasticity of complementarity (MEC), and the Symmetric elasticity of complementarity (SEC). The formulas are provided here for convenience:

$$AEC_{kl} = \frac{D \cdot D_{kl}}{D_k \cdot D_l}, \quad (23)$$

$$MEC_{kl} = \frac{D_{kl} \cdot X_l}{D_k} - \frac{D_{ll} \cdot X_l}{D_l}, \quad (24)$$

and

$$SEC_{kl} = \frac{-\frac{D_{kk}}{D_k D_k} + 2 \frac{D_{kl}}{D_k D_l} - \frac{D_{ll}}{D_l D_l}}{\frac{1}{D_k X_k} + \frac{1}{D_l X_l}}, \quad (25)$$

$\forall k, l = 1, \dots, K$. D_k (D_{kk}) and D_l (D_{ll}) are the first (second) partial derivatives of the IDF with respect to the k th and l th input, respectively; and D_{kl} are the cross partial derivatives with respect to the k th and l th input. Here both the AEC and the SEC are symmetric: they give the same elasticity estimate no matter what input causes a change. The MEC is not symmetric for the arguments made in Blackorby and Russell (1981). If the MEC > 0 , then the two inputs are complements in the Morishima sense. Similar interpretation applies to the AEC and the SEC. For comparison purposes, in the application section, the elasticities are calculated from the IDF both with and without the theoretical restrictions.

3 Application

As an empirical illustration of the proposed methodology, we use a cross-sectional data set of 3,249 active forest owners (i.e., owners who harvest trees) for the year 2003 compiled by Statistics Norway. According to Statistics Norway,¹¹ the value added in Norwegian forestry was estimated at Norwegian Krone (NOK) 5.4 billion in 2011. Timber sale is the

¹¹ See <http://www.ssb.no/english/subjects/10/04/20/>.

largest component in Norwegian forestry. From 2001 to 2011, 36 per cent of the forest properties sold timber. In addition, forest owners can also earn income from selling hunting and fishing rights, leasing out sites and renting out cabins. The Ministry of Agriculture and Food of Norway (2007) reported that approximately 88 per cent of the forest area is privately owned, and the majority of the forest holdings are farm and family forests. During the last 80 years, timber stock increased because the annual timber growth has been considerably faster than the annual harvest.

Since this data set has been used in Lien et al (2007) in which a detailed description of the sampling method is available, a brief description of the data is given as follows. The output variable (Y) consists of annual timber sales from the forest, measured in cubic meters. The labor input variable (X_1) is the sum of hours worked by contractors and hours worked by the owner, his family or hired labor. The land input variable (X_2) measures the forest area to be cut in hectares. The capital input variable (X_3) is the amount of timber stock that can be cut without affecting future harvesting. Table 1 presents summary statistics in the sample.

The estimation results are given in Tables 2-3 and Figures 1-4. Table 2 reports the estimated bandwidth vector and shows percentages of violations of the monotonicity properties of the IDF. Table 3 presents the Kolmogorov-Smirnov testing results for equality of distributions between the information estimated from the unconstrained and the constrained models. Figure 1 reports the histogram of observation-specific weights for imposing the monotonicity constraints. Figures 2-4 plot the kernel densities of the gradient and elasticity estimates under the unconstrained and the constrained models.

It can be seen from Table 2 that, the bandwidth estimate for each regressor is small enough (i.e., less than twice the standard deviation of the corresponding regressor) to indicate nonlinearity of the regression function, hence the appropriateness of the nonparametric approach. Using these bandwidth estimates, although nearly no violation of economic theory occurs for the gradients of $\ln X_1$ and $\ln Y$, there are 18.25% and 7.08% of violations for the gradients of $\ln X_2$ and $\ln X_3$, respectively.¹² This suggests that it should not be trivial to impose the economic constraints of (15) and (16).

The constraint weighted bootstrapping (CWB) method is then used to impose these constraints. Figure 1 plots the distribution of p_i^* , the optimal weight for each observation suggested by CWB. It can be seen that most observations share similar weights, and these optimal weights are quite close to the uniform weights, $1/n = 1/3249 \approx 3 \times 10^{-4}$. After the dependent variable is transformed by these weights, we can then use them to estimate the gradients and the elasticities under the constraints.

Figure 2 shows the kernel density estimates for the unconstrained and constrained distributions of the four gradients, i.e., $\partial \ln D / \partial \ln X_k$, $\forall k = 1, 2, 3$, and $\partial \ln D / \partial \ln Y$, on which for each observation, a non-negativity constraint is imposed on the first three gradients, and a non-positivity constraint is imposed on the last one. A vertical line is drawn at zero and the shaded area highlights where the violations occur. The densities for the constrained gradients are plotted using the Silverman reflection method for boundary correction such that the estimated densities integrate to one. It can be seen that there are masses near zero with the constrained model for the gradients with many violations. Although very few violations are observed for the gradient of $\ln X_1$, the unconstrained and constrained distributions of it are not close

¹² We also checked whether the estimated IDF is concave in all the inputs: before the monotonicity constraints are imposed, there are 2,080 out of the 3,249 observations, or about 64.02% of the observations that satisfy the input concavity condition, and after the monotonicity constraints are imposed, there are 2,137, or about 65.77% of the observations that satisfy the input concavity condition.

to each other. This is because the gradient of the first input is essentially a linear combination of the gradients of the other two inputs, whose non-trivial amount of violations affect the constrained gradient of the first input.

We also plot the kernel density estimates of the elasticities under the unconstrained and constrained models, as Figure 3 and 4 illustrate. It can be seen that, in most cases, the constrained elasticities have smaller variation than their unconstrained counterparts. This suggests that estimating an IDF satisfying its properties may improve the efficiency of the elasticity estimates from the IDF. A vertical line is drawn at zero for a better view of the percentage of observations whose inputs are substitutes/complements. For example, if the MEC between any two inputs is greater (less) than zero, then these inputs are Morishima complements (substitutes). However, for both Figures, the elasticity estimates from the unconstrained and constrained models seem to be quite close to each other. To convince the reader that significant differences exist between the distributions, Table 3 reports the p -values from the Kolmogorov-Smirnov test for equality of distributions for the gradients and the elasticities estimated from the unconstrained and constrained models. It can be seen that the null of equality of distributions is rejected at the 5% level in all cases except the AEC estimates between labor and capital.

4 Discussion

4.1 Choice of numeraire input

When it comes to estimating the IDF with some parametric functional form, e.g., Cobb-Douglas or Translog, the choice of the normalizing input does not affect the results. However, this is generally not the case for nonparametric estimation of the IDF.¹³ The empirical example chooses labor input as the numeraire input when estimating the IDF. This assumes that the labor input is endogenous. In order to see how the results change when other inputs (i.e., land or capital) are used for normalization, we simply report the unconstrained and the constrained gradient estimates in Figures 5 and 6 in Appendix C, which uses land and capital input as the numeraire, respectively. The elasticity plots are omitted to save space. Although it is recommended that researchers choose an input that is endogenous *a priori*, the question of how to find the most appropriate numeraire may be answered in future research.

4.2 Extensions to other representations of technology

The estimation procedure in section 2 can be easily extended to other specifications of technology. We provide two examples here: a production and a cost function.

The production function can be written as $y = B \cdot F(X)$, where y is a scalar output, B is the productivity parameter, $F(\cdot)$ is an unknown function, and X is an input vector. Applying log transformation similar to (6) gives an estimable production function: $\ln y = f(\ln X) + u$, where $u = \ln B$ is the error term. The unknown function $f(\cdot)$ can be estimated nonparametrically via kernel methods. One can also impose some desirable constraints on $f(\cdot)$, e.g., $\partial \ln y / \partial \ln X \geq 0$ as a non-negative marginal product constraint.

¹³ The author would like to thank an anonymous referee for this observation.

For the cost function, it can be written as $C = \lambda C(W, Y)$, where C is the total cost, λ is the productivity parameter, $C(\cdot)$ is an unknown function, and W and Y are input price and output vectors, respectively. Applying the log transformation and imposing the homogeneity of degree one restriction in input price gives an estimable cost function: $\ln \tilde{C} = c(\ln \tilde{W}, \ln Y) + u$, where \tilde{C} is the total cost divided by a numeraire input price, \tilde{W} is the input price vector divided by the numeraire input price, and $u = \ln \lambda$ is the error term. The cost function $c(\cdot)$ can be estimated using kernel methods, and some necessary regularity constraints can be imposed on it such that (1) cost shares are constrained between zero and one: $0 \leq \partial \ln \tilde{C} / \partial \ln \tilde{W} \leq 1$, and (2) marginal cost is non-negative: $\partial \ln \tilde{C} / \partial \ln Y \geq 0$.

After the regression functions are estimated under the constraints, different elasticities can then be calculated subject to the specification of choice and data availability. See Stern (2011) for a classification scheme of different definitions of elasticities based on primal and dual representations of technology.

4.3 Possibly endogenous regressors

It is well known that estimation of production/cost/distance functions may be subject to the endogeneity problem that causes estimation results to be biased and inconsistent. Unfortunately, the nonparametric instrumental variable (IV) estimation is a quite young field - see Su and Ullah (2008) for a three-step estimation procedure for nonparametric simultaneous equations models via kernel methods, and Newey and Powell (2003) for IV estimation via series approximation, among others. It is unclear whether the CWB procedure can be seamlessly applied to the nonparametric structural models, which may be saved for future research.

5 Conclusion

This paper uses econometric methods to estimate an input distance function (IDF) without functional form assumptions, and imposes economic properties of the IDF on the estimated regression function via constraint weighted bootstrapping (CWB). As a by-product, the first, second, and cross partial analytical derivatives of the estimated IDF are derived, and thus various elasticities can be computed. Applying the proposed method to a cross-section of Norwegian forest owners, we find that imposing the constraints eliminates the problem of economic violations in empirical work, and therefore policy implications may be more reliable. The proposed method can be extended to other representations of technology in a straightforward manner, and this opens the door for further empirical work to estimate models subject to economic theory. As a future research topic, more work should be done on the unification of CWB and the nonparametric structural modeling approach.

Acknowledgements The author would like to thank Gudbrand Lien for providing the data set, Daniel J. Henderson, Subal C. Kumbhakar, Christopher F. Parmeter, attendees at the 2011 Econometric Society North American Winter Meeting, and the two anonymous referees for useful comments. The author is responsible for any remaining errors.

Table 1 Summary Statistics of the Variables

Symbol	Variable Name	Variable Description	Mean	Sd.	Min.	Max.
Y	Output(m^3)	Harvesting level	997.4	2621.2	2	46070
X_1	Labor(hours)	Working hours	58.04	141.2	0.126	2632
X_2	Land(hectares)	Forest area cut	5.97	14.8	0.012	229
X_3	Capital(NOK)	Timber stock	313740	821481	1476	16190000

Table 2 Bandwidths and Percentages of Violations

Bandwidths	-	$\ln \tilde{X}_2$	$\ln \tilde{X}_3$	$\ln Y$
	-	0.2869	0.5608	0.1473
Violations	$\partial \ln D / \partial \ln X_1 < 0$	$\partial \ln D / \partial \ln X_2 < 0$	$\partial \ln D / \partial \ln X_3 < 0$	$\partial \ln D / \partial \ln Y > 0$
Percentages	0.12%	18.25%	7.08%	0.31%

1. The subscript indices are: 1 =Labor, 2 =Land, and 3 =Capital.

Fig. 1 Observation-specific Weights

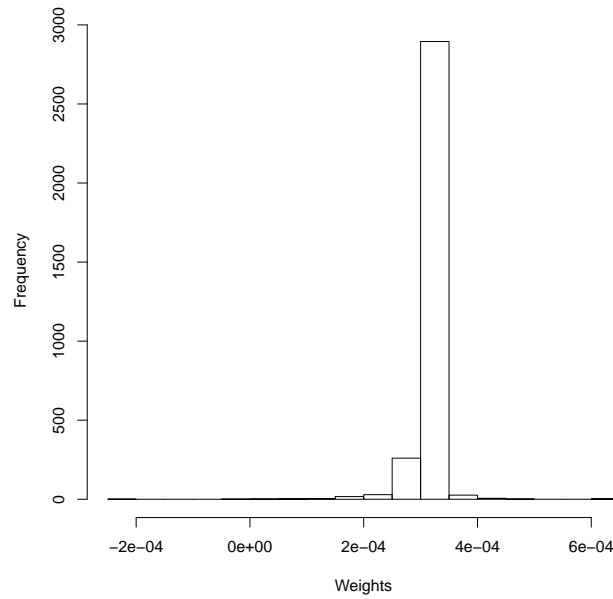


Fig. 2 Kernel Density Plots of the Gradients of the IDF: Unconstrained versus Constrained

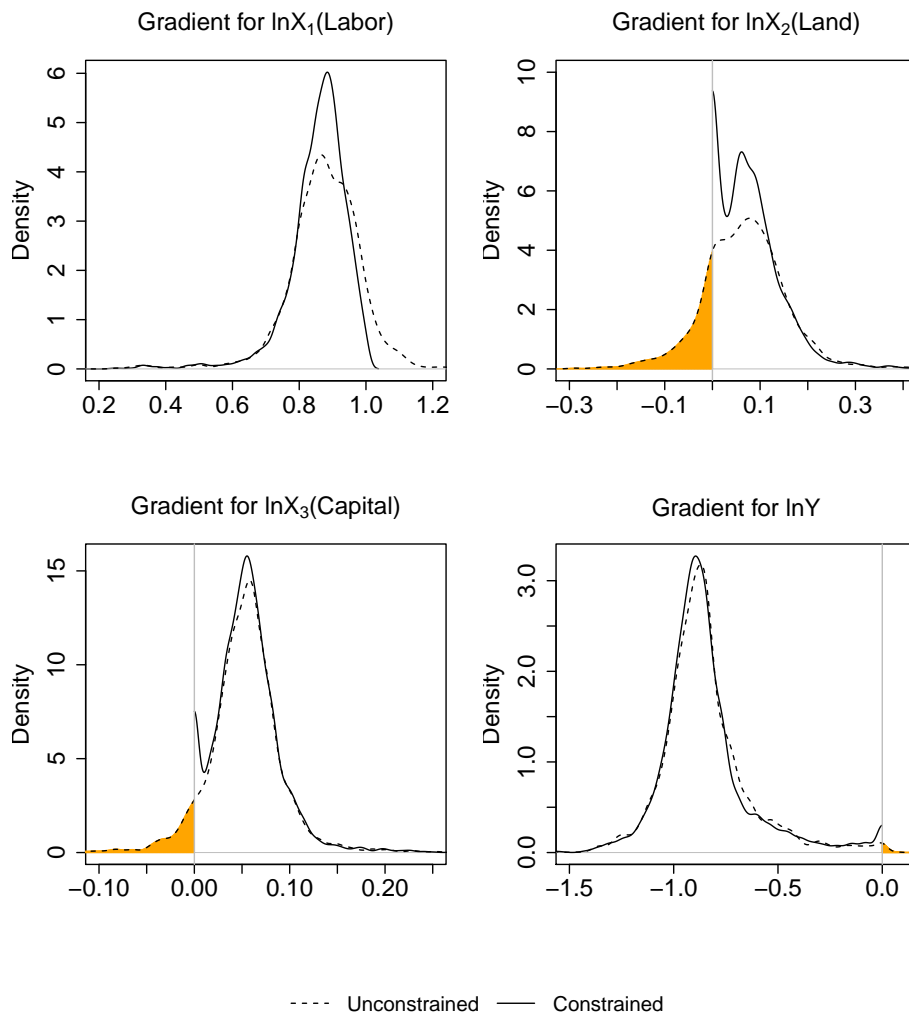


Table 3 Testing for Equality of Distributions from the Unconstrained and Constrained Models

Gradients	$\partial \ln D / \partial \ln X_1$	$\partial \ln D / \partial \ln X_2$	$\partial \ln D / \partial \ln X_3$	$\partial \ln D / \partial \ln Y$	-	-
<i>p</i> -values	0.0000	0.0000	0.0000	0.0221	-	-
Elasticities						
	MEC_{12}	MEC_{13}	MEC_{23}	MEC_{21}	MEC_{31}	MEC_{32}
<i>p</i> -values	0.0000	0.0163	0.0002	0.0000	0.0129	0.0000
	AEC_{12}	AEC_{13}	AEC_{23}	SEC_{12}	SEC_{13}	SEC_{23}
<i>p</i> -values	0.0000	0.0552	0.0044	0.0000	0.0102	0.0000

1. The subscript indices are: 1 =Labor, 2 =Land, and 3 =Capital.

2. The *p*-values are obtained from the Kolmogorov-Smirnov test for equality of distributions against the two-sided alternatives.

Fig. 3 Kernel Density Plots of the Nonparametric MEC: Unconstrained versus Constrained

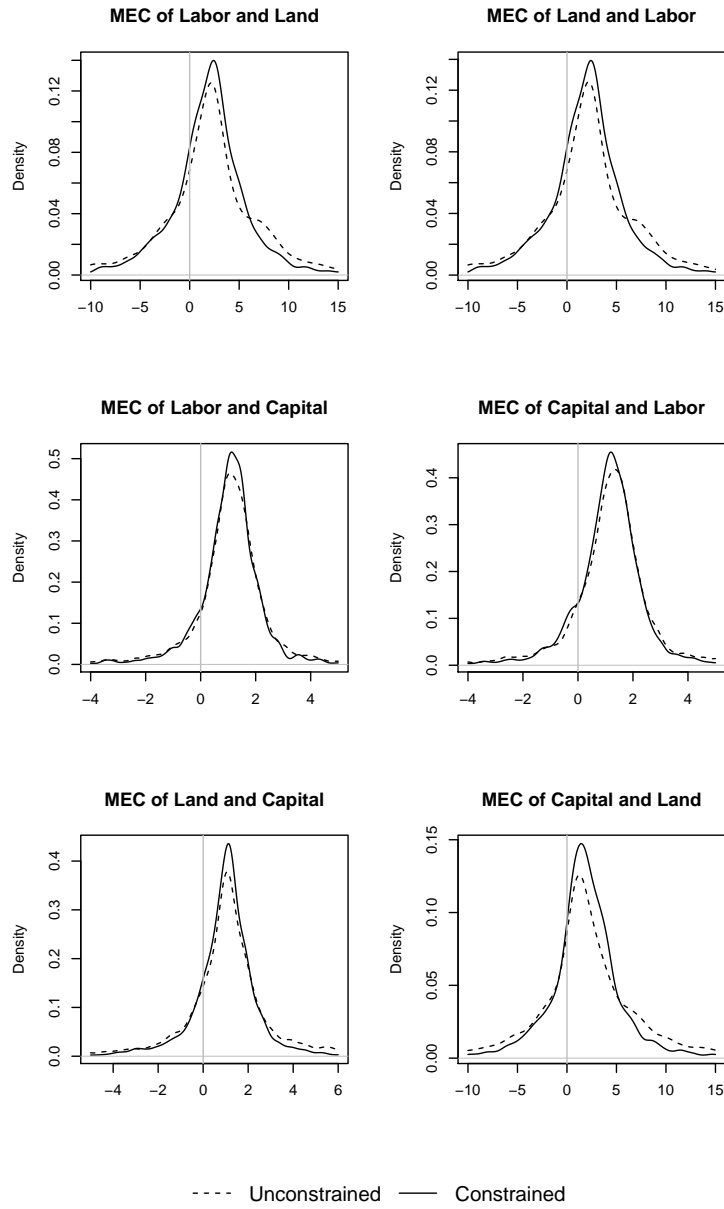
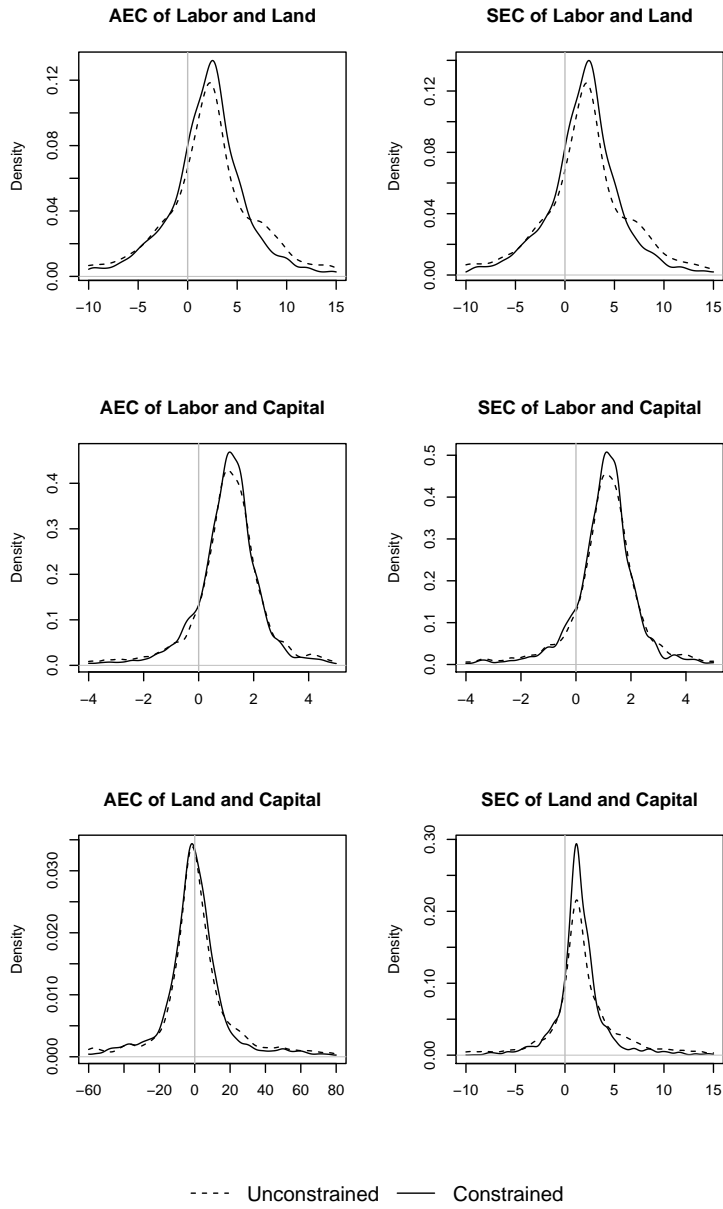


Fig. 4 Kernel Density Plots of the Nonparametric AEC and SEC: Unconstrained versus Constrained



References

- Berndt E, Christensen L (1973) The translog function and the substitution of equipment, structures, and labor in U.S. manufacturing 1929-68. *Journal of Econometrics* 1(1):81-113
- Berwin A, Weingessel A (2011) quadprog: Functions to solve Quadratic Programming Problems.
- Blackorby C, Russell R (1981) The Morishima elasticity of substitution: Symmetry, constancy, separability, and its relationship to the Hicks and Allen elasticities. *Review of Economic Studies* 48:147-158
- Coelli T, Perelman S (1999) A comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operations Research* 117:326-339
- Cuesta R, Lovell C, Zofio J (2009) Environmental efficiency measurement with translog distance functions: A parametric approach. *Ecological Economics* 68:2232-2242
- Du P, Parmeter C, Racine J (2013) Constrained nonparametric kernel regression: Estimation and inference. *Statistica Sinica*, forthcoming
- Färe R, Primont D (1995) *Multi-output Production and Duality: Theory and Applications*. Kluwer Academic Publishers, Boston
- Färe R, Grosskopf S, Lovell C (1985) *The Measurement of Efficiency of Production*. Kluwer Academic Publishers, Boston
- Färe R, Grosskopf S, Lovell C (1994) *Production Frontiers*. Cambridge University Press, Cambridge
- Grosskopf S, Hayes K, Taylor L, Weblar W (1997) Budget constrained frontier measures of fiscal equality and efficiency in schooling. *Review of Economics and Statistics* 79:116-124
- Hall P, Huang H (2001) Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics* 29(3):624-647
- Hayfield T, Racine J (2008) Nonparametric econometrics: The np package. *Journal of Statistical Software* 27(5)
- Henderson D (2009) A non-parametric examination of capital-skill complementarity. *Oxford Bulletin of Economics and Statistics* 71:519-538
- Henningsson A, Henning C (2009) Imposing regional monotonicity on translog stochastic production frontiers with a simple three-step procedure. *Journal of Productivity Analysis* 32:217-229
- Heshmati A, Kumbhakar S, Sun K (2013) Estimation of productivity in Korean electric power plants: A semiparametric smooth coefficient model. Working paper
- Kumbhakar S, Lovell C (2000) *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge
- Li Q, Racine J (2004) Cross-validated local linear nonparametric regression. *Statistica Sinica* 14:485-512
- Li Q, Racine J (2006) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton
- Lien G, Størdal S, Baardsen S (2007) Technical efficiency in timber production and effects of other income sources. *Small-scale Forestry* 6:65-78
- Lovell C, Richardson S, Travers P, Wood L (1994) Resources and functionings: A new view of inequality in Australia, Springer-Verlag, Berlin, pp 787-807. *Models and Measurement of Welfare and Inequality*
- Newey W, Powell J (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71(5):1565-1578
- Parmeter C, Sun K, Henderson D, Kumbhakar S (2013) Estimation and inference under economic restrictions. *Journal of Productivity Analysis*, forthcoming
- R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria
- Rambaldi A, Doran H (1997) Applying linear time-varying constraints to econometric models: With an application to demand systems. *Journal of Econometrics* 79(1):83-95
- Ray S (2003) Measuring scale efficiency from the translog multi-input, multi-output distance function. Department of Economics, University of Connecticut, no. 2003-25, Working papers
- Ryan D, Wales T (2000) Imposing local concavity in the translog and generalized Leontief cost functions. *Economics Letters* 67:253-260
- Shephard R (1953) *Cost and Production Functions*. Princeton University Press, Princeton
- Shephard R (1970) *Theory of Cost and Production Functions*. Princeton University Press, Princeton
- Stern D (2011) Elasticities of substitution and complementarity. *Journal of Productivity Analysis* 36(1):79-89
- Su L, Ullah A (2008) Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics* 144(1):193-218
- Terrell D (1996) Incorporating monotonicity and concavity conditions in flexible functional forms. *Journal of Applied Econometrics* 11(2):179-194
- The Ministry of Agriculture and Food of Norway (2007) *Norwegian forests: Policy and resources*. Address: Teatergata 9 (R6), Postbox. 8007 Dep. 0030 Oslo, Norway

Appendix A

In order to obtain the derivative of the distance function with respect to each input in level form, we start from the equation in log form:

$$\ln D(X, Y) = \ln X_1 + m(\ln \tilde{X}_2, \dots, \ln \tilde{X}_K, \ln Y_1, \dots, \ln Y_Q)$$

where $\ln \tilde{X}_k = \ln X_k - \ln X_1$, $\forall k = 2, \dots, K$.

$$\frac{\partial \ln D}{\partial \ln X_1} = 1 - \sum_{k=2}^K \frac{\partial m}{\partial \ln \tilde{X}_k}$$

$$\frac{\partial \ln D}{\partial \ln X_k} = \frac{\partial m}{\partial \ln \tilde{X}_k}, \quad \forall k = 2, \dots, K$$

$$\frac{\partial^2 \ln D}{\partial \ln X_1^2} = - \sum_{k=2}^K \frac{\partial^2 m}{\partial \ln \tilde{X}_k^2} \cdot \frac{\partial \ln \tilde{X}_k}{\partial \ln X_1} = \sum_{k=2}^K \frac{\partial^2 m}{\partial \ln \tilde{X}_k^2}$$

$$\frac{\partial^2 \ln D}{\partial \ln X_k^2} = \frac{\partial^2 m}{\partial \ln \tilde{X}_k^2}, \quad \forall k = 2, \dots, K$$

$$\frac{\partial^2 \ln D}{\partial \ln X_1 \partial \ln X_k} = - \frac{\partial^2 m}{\partial \ln \tilde{X}_k^2}, \quad \forall k = 2, \dots, K$$

$$\frac{\partial^2 \ln D}{\partial \ln X_k \partial \ln X_l} = \frac{\partial^2 m}{\partial \ln \tilde{X}_k \partial \ln \tilde{X}_l}, \quad \forall k, l = 2, \dots, K \text{ and } k \neq l$$

Once we obtain the derivatives in log form, it would be straightforward to recover the derivatives in level form.

$$D_k = \frac{\partial D}{\partial X_k} = \frac{\partial \ln D}{\partial \ln X_k} \cdot \frac{D}{X_k}, \quad \forall k = 1, \dots, K$$

$$D_{kk} = \frac{\partial^2 D}{\partial X_k^2} = \frac{\partial^2 \ln D}{\partial \ln X_k^2} \cdot \frac{D}{X_k^2} + \frac{1}{X_k} \left(D_k - \frac{1}{X_k} D \right) \frac{\partial \ln D}{\partial \ln X_k}, \quad \forall k = 1, \dots, K$$

$$D_{kl} = \frac{\partial^2 D}{\partial X_k \partial X_l} = \frac{\partial^2 \ln D}{\partial \ln X_k \partial \ln X_l} \cdot \frac{D}{X_k X_l} + D_l \cdot \frac{1}{X_k} \cdot \frac{\partial \ln D}{\partial \ln X_k}, \quad \forall k, l = 1, \dots, K \text{ and } k \neq l$$

Appendix B

This appendix derives the first, second, and cross partial analytical derivatives of $\hat{m}(z)$ with respect to the l th continuous variable z_l .

$$\hat{m}(z) = \sum_{i=1}^n A_i(z) \mathcal{Y}_i$$

where

$$A_i(z) = \frac{K\left(\frac{z_i - z}{h}\right)}{\sum_{i=1}^n K\left(\frac{z_i - z}{h}\right)}$$

where $\mathcal{Y}_i = -\ln X_{1i}$, and $K(\cdot)$ is a product kernel function:

$$K(\cdot) = \prod_{s=1}^q K\left(\frac{z_{si} - z_s}{h_s}\right)$$

For the l th continuous variable z_l ,

$$K\left(\frac{z_{li} - z_l}{h_l}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{z_{li} - z_l}{h_l}\right)^2\right)$$

where h_l denotes the bandwidth for z_l .

$$\begin{aligned} \frac{\partial^r \hat{m}(z)}{\partial z_l^r} &= \sum_{i=1}^n \frac{\partial^r A_i(z)}{\partial z_l^r} \mathcal{Y}_i, \quad \forall r = 1, 2 \\ \frac{\partial^2 \hat{m}(z)}{\partial z_l \partial z_k} &= \sum_{i=1}^n \frac{\partial^2 A_i(z)}{\partial z_l \partial z_k} \mathcal{Y}_i, \quad \forall l \neq k. \end{aligned}$$

Therefore, the derivatives of $\hat{m}(z)$ with respect to continuous z variables can be expressed in terms of the derivatives of $A_i(z)$ with respect to these variables. Specifically,

$$\begin{aligned} \frac{\partial A_i(z)}{\partial z_l} &= \frac{T(\cdot)}{(\sum_{i=1}^n K(\cdot))^2} \\ \frac{\partial^2 A_i(z)}{\partial z_l^2} &= \frac{\frac{\partial T(\cdot)}{\partial z_l} (\sum_{i=1}^n K(\cdot))^2 - 2T(\cdot) (\sum_{i=1}^n K(\cdot)) \left(\sum_{i=1}^n \frac{\partial K(\cdot)}{\partial z_l}\right)}{(\sum_{i=1}^n K(\cdot))^4} \\ \frac{\partial^2 A_i(z)}{\partial z_l \partial z_k} &= \frac{\frac{\partial T(\cdot)}{\partial z_k} (\sum_{i=1}^n K(\cdot))^2 - 2T(\cdot) (\sum_{i=1}^n K(\cdot)) \left(\sum_{i=1}^n \frac{\partial K(\cdot)}{\partial z_k}\right)}{(\sum_{i=1}^n K(\cdot))^4} \end{aligned}$$

where

$$\begin{aligned} T(\cdot) &= \frac{\partial K(\cdot)}{\partial z_l} \sum_{i=1}^n K(\cdot) - K(\cdot) \sum_{i=1}^n \frac{\partial K(\cdot)}{\partial z_l} \\ \frac{\partial T(\cdot)}{\partial z_l} &= \frac{\partial^2 K(\cdot)}{\partial z_l^2} \sum_{i=1}^n K(\cdot) - K(\cdot) \sum_{i=1}^n \frac{\partial^2 K(\cdot)}{\partial z_l^2} \\ \frac{\partial T(\cdot)}{\partial z_k} &= \frac{\partial^2 K(\cdot)}{\partial z_l \partial z_k} \sum_{i=1}^n K(\cdot) - K(\cdot) \sum_{i=1}^n \frac{\partial^2 K(\cdot)}{\partial z_l \partial z_k} + \frac{\partial K(\cdot)}{\partial z_l} \sum_{i=1}^n \frac{\partial K(\cdot)}{\partial z_k} - \frac{\partial K(\cdot)}{\partial z_k} \sum_{i=1}^n \frac{\partial K(\cdot)}{\partial z_l} \end{aligned}$$

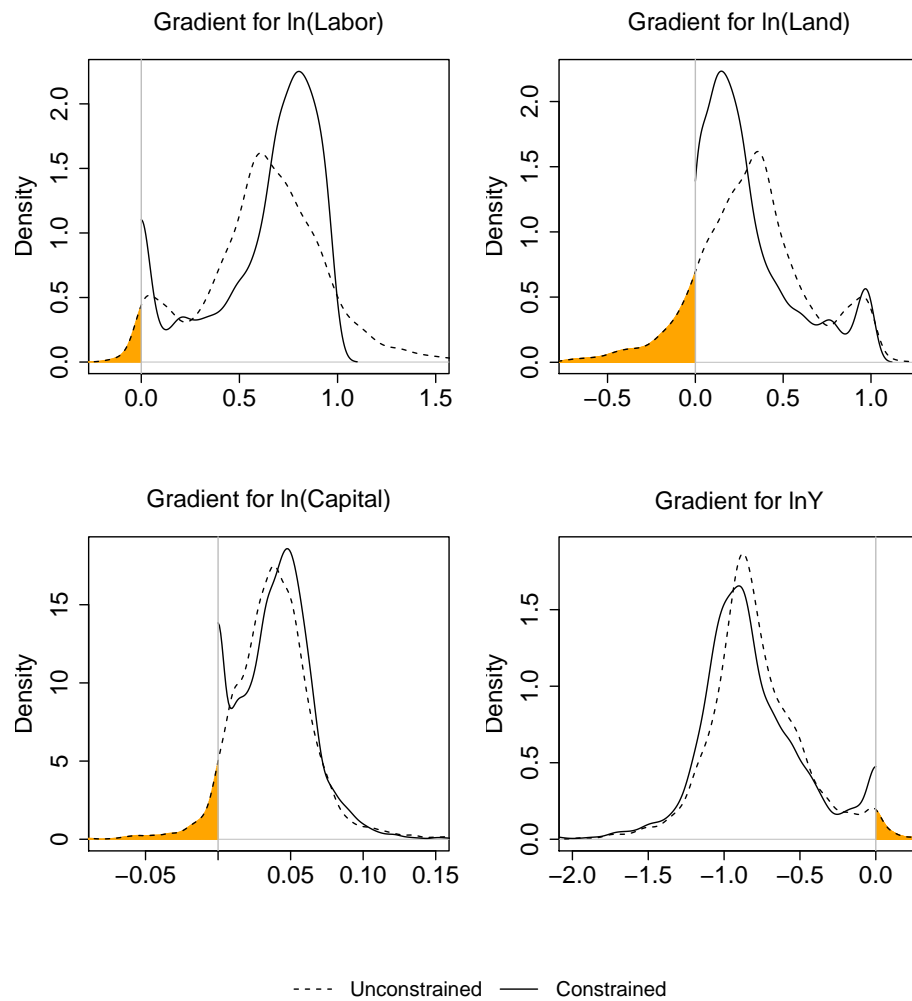
We can see the derivatives of $A_i(z)$ are functions of $T(\cdot)$ and its derivatives. In order to calculate $T(\cdot)$ and its derivatives, we need to calculate the first, second, and cross partial derivatives of $K(\cdot)$ with respect to the continuous variables:

$$\begin{aligned} \frac{\partial K(\cdot)}{\partial z_l} &= \left(\frac{z_{li} - z_l}{h_l^2}\right) K(\cdot) \\ \frac{\partial^2 K(\cdot)}{\partial z_l^2} &= \left[\frac{(z_{li} - z_l)^2 - h_l^2}{h_l^4}\right] K(\cdot) \\ \frac{\partial^2 K(\cdot)}{\partial z_l \partial z_k} &= \left(\frac{z_{li} - z_l}{h_l^2}\right) \left(\frac{z_{ki} - z_k}{h_k^2}\right) K(\cdot), \quad \forall l \neq k. \end{aligned}$$

Appendix C

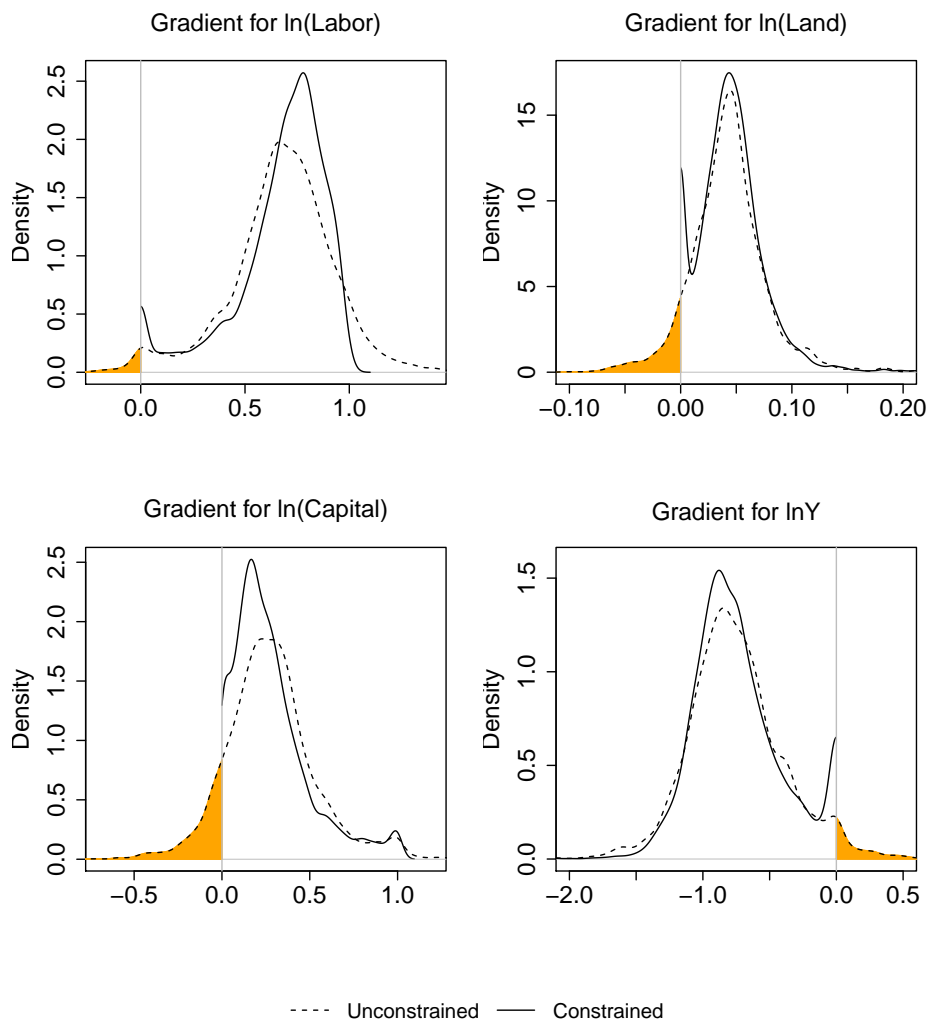
This appendix contains kernel density plots of the gradients of the IDF when alternative inputs are chosen as the numeraire input.

Fig. 5 Kernel Density Plots of the Gradients of the IDF: Unconstrained versus Constrained*



* Land input is chosen as the numeraire input.

Fig. 6 Kernel Density Plots of the Gradients of the IDF: Unconstrained versus Constrained*



* Capital input is chosen as the numeraire input.