Direction Selection in Stochastic Directional Distance Functions

Kevin Layer[a], Andrew L. Johnson[a,b,*], Robin C. Sickles[c], Gary D. Ferrier[d]

[a]*Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA.*
[b]*School of Information Science and Technology, Osaka University, Suita, Japan.*
[c]*Department of Economics, Rice University, Houston, TX, USA.*
[d]*Department of Economics, University of Arkansas, Fayetteville, AR, USA.*

Abstract

Researchers rely on the distance function to model multiple product production using multiple inputs. A stochastic directional distance function (SDDF) allows for noise in potentially all input and output variables, yet when estimated, the direction selected will affect the functional estimates because deviations from the estimated function are minimized in the specified direction. Specifically, the parameters of the parametric SDDF are point identified when the direction is specified, we show that the parameters of the parametric SDDF are set identified when multiple directions are considered. Further, the set of identified parameters can be narrowed via data-driven approaches to restrict the directions considered. We demonstrate a similar narrowing of the identified parameter set for a shape constrained nonparametric method, where the shape constraints impose standard features of a cost function such as monotonicity and convexity.

Our Monte Carlo simulation studies reveal significant improvements, as measured by out of sample radial mean squared error, in functional estimates when we use a directional distance function with an appropriately selected direction and the errors are uncorrelated across variables. We show that these benefits increase as the correlation in error terms across variables increase. This correlation is a type of endogeneity that is common in production settings. From our Monte Carlo simulations we conclude that selecting a direction that is approximately orthogonal to the estimated function in the central region of the data gives significantly better estimates relative to the directions commonly used in the literature. For practitioners, our results imply that selecting a direction vector that has non-zero components for all variables that may have measurement error provides a significant improvement in the estimator's performance. We illustrate these results using cost and production data from three random samples of approximately 500 US hospitals operating in 2007, 2008, and 2009, respectively, and find that the shape constrained nonparametric methods provide a significant increase in flexibility over second order local approximation parametric methods.

Keywords: Nonparametric regression, Shape Constraints, Data Envelopment Analysis, Hospital production.

*Communicates regarding this paper can be sent to Andrew L. Johnson, ajohnson@tamu.edu

1. Introduction

The focus of this paper is direction selection in stochastic directional distance functions (SDDF).[1] While the DDF is typically used to measure efficiency, in this paper we use a nonparametric shape constrained SDDF to model the conditional-mean behavior of production. The stochastic distance functions (SDF) was introduced by Lovell et al. (1994), and used in a series of early empirical studies by Coelli and Perelman (1999, 2000) and Sickles et al. (2002). The parameters of a parametric distance function are point identified; however, if the direction in the DDF is not specified, then the parameters of a parametric DDF are set identified.[2] A set of axiomatic properties related to production and cost functions, such as monotonicity and convexity in the case of a cost function, are well established in the production literature (Shephard (1970), Chambers (1988)). Although the stochastic distance function literature acknowledges the axiomatic properties necessary for duality, it does not impose them globally. Instead, authors typically impose them only on a particular point in the data (e.g., Atkinson et al. (2003)). Recognizing these issues, we provide an axiomatic nonparametric estimator of the SDDF and a method to restrict the pool of the directions to choose from for the SDDF, thereby reducing the size of the set identified parameter set. Most empirical studies that use establishment or hospital level data to estimate production or cost functions either assume a specific parametric form or ignore noise, or both ((Hollingsworth, 2003)). In contrast, we use an axiomatic nonparametric SDDF estimator and the proposed method to determine a set of acceptable directions to estimate a cost function that maintains global axiomatic properties for the US hospital industry. Furthermore, we demonstrate the importance of global axiomatic properties for the estimation of most productive scale size and marginal costs.

A few papers have attempted to implement the directional distance function in a stochastic setting (see, for example, Färe et al. (2005), Färe et al. (2010), and Färe and Vardanyan (2016)). The latter two papers discuss the challenges of selecting a parametric functional form that does not violate the axioms typically assumed in production economics. Based on their observations, Färe and Vardanyan (2016) use a quadratic functional specification.[3] Yet several papers show also a loss of flexibility in parametric functional forms, such as the translog or the quadratic functional form, when shape constraints are imposed (e.g., Diewert and Wales (1987)). Another part of the implementation, the selection of the direction vector in the SDDF has been discussed in Färe et al. (2016) and Atkinson and Tsionas (2016), among others. These papers focus on selecting the direction corresponding to a particular interpretation of the inefficiency measure, based on the

[1] Here we use the term stochastic in reference to a model with a noise term.

[2] Let ϕ be what is known (e.g., via assumptions and restrictions) about the data generating process (DGP). Let θ represent the parameters to be identified, let Θ denote all possible values of θ, and let θ_0 be the true but unknown value of θ. Then the vector θ of unknown parameters is point identified if it is uniquely determined from ϕ. However, θ is set identified if some of the possible values of θ are observationally equivalent to θ_0 (Lewbel (2018), forthcoming).

[3] As Kuosmanen and Johnson (2017) note, the translog function used for multi-output production cannot satisfy the standard assumptions for the production technology T globally for any parameter values. The quadratic functional form does not have this shortcoming.

distance to the economically efficient point. In contrast, we consider Kuosmanen and Johnson (2017)'s multi-step efficiency analysis and focus on the first step, estimating a conditional mean function. Our goal is to select the direction that best recovers the underlying technology while acknowledging that the data generation process is very likely to introduce noise into potentially all variables.[4]

To model multi-production production, Kuosmanen and Johnson (2017) have proposed the use of axiomatic nonparametric methods to estimate the SDDF which they name Directional Convex Nonparametric Least Squares(CNLS-d), a type of sieve estimator. Their methods have the benefits of relaxing standard functional form assumptions for production, cost, or distance functions, but also improve the interpretability and finite sample efficiency over nonparametric methods such as kernel regression (Yagi et al. (2018)). A variety of models can be interpreted as special cases of Kuosmanen and Johnson (2017). Among these special cases are a set of models that specify the direction (e.g., Johnson and Kuosmanen (2011), Kuosmanen and Kortelainen (2012)). All CNLS models are sieve estimators and fall into the category of partially identified or set identified estimators discussed in Manski (2003) and Tamer (2010). The guidance our paper provides in selecting a direction will reduce the size of the set identified for CNLS-d and other DDF estimators with flexible direction specifications.

Much of the production function literature concerns endogeneity issues, for example Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015). These methods are often referred to as proxy variable approaches. The argument for endogeneity is typically that decisions regarding variable inputs such as labor are made with some knowledge of the factors included in the unobserved residuals. Recently, these methods have been reinterpreted as instrumental variable approaches (Wooldridge (2009)), or control function approaches (Ackerberg et al. (2015)). Unfortunately, the assumptions on the particular timing of input decisions is not innocuous. Indeed every firm must adjust its inputs in exactly the same way, otherwise the moment restrictions needed for point identification are violated. For an alternative in the stochastic frontier setting, see Kutlu (2018).

Kuosmanen and Johnson (2017) show a production function estimated using a stochastic distance function under a constant returns-to-scale assumption is robust to endogeneity issues because the normalization by one of the inputs or outputs causes the errors-in-variables to cancel each other. In this paper, we consider the more general case of a convex technology, that does not necessarily satisfy constant returns-to-scale, and show that when errors across variables are highly correlated, a specific type of endogeneity, the SDDF improves estimation performance significantly over the typical alternative of ignoring the endogeneity.

[4]For researchers interested in productivity measurement and productivity variation (e.g., Syverson (2011)), the results from this paper can be used directly. For authors interested in efficiency analysis, they can use the insights from this paper to improve the estimates from the first stage of Kuosmanen and Johnson's (2017) three-step procedure where efficiency is estimated in the third step.

When considering alternative directions in the DDF, we show that the direction that performs the best is often related to the particular performance measure used. We use an out-of-sample mean squared error (MSE) that is measured radially to address this issue. This measure is motivated by the results of our Monte Carlo simulations and is natural for a function that satisfies monotonicity and convexity, assuring the true function and the estimated function are close in the areas were most data are observed.

We analyze US hospital data and characterize the most productive scale size and marginal costs for the US hospital sector. We demonstrate that out-of-sample MSE is reduced significantly by relaxing parametric functional form restrictions. We also observe the advantage of imposing axioms that allow the estimated function to still be interpretable. Concerning the direction selection, we find, for this dataset, that the exact direction selected is not so critical in terms of MSE performance.

The remainder of this paper is organized as follows. Section 2 introduces the statistical model and the production model. Section 3 describes the estimators used for the analysis. Section 4 outlines our reasons for the MSE measure we propose. Section 5 highlights the importance of the direction selection through Monte Carlo experiments. Section 6 describes our direction selection method. Section 7 demonstrates the benefits of using non-parametric shape-constrained estimators with an appropriately selected direction for US hospital data. Section 8 concludes.

2. Models

2.1. Statistical Model

We consider a statistical model that allows for measurement error in potentially all of the input and output variables. Let $\tilde{\boldsymbol{x}}_i \in \boldsymbol{X} \subset \mathbb{R}^d_+, d \geq 1$, be a vector of random input variables of length d and $\tilde{\boldsymbol{y}}_i \in \boldsymbol{Y} \subset \mathbb{R}^Q_+$, $Q \geq 1$, be a vector of random output variables of length Q, where i indexes observations. Let $\boldsymbol{\epsilon}^x_i \in \mathbb{R}^d$, $d \geq 1$, be a vector of random error variables of length d and $\boldsymbol{\epsilon}^y_i \in \mathbb{R}^Q$, $Q \geq 1$, be a vector of random error variables of length Q. One way of modeling the errors-in-variable (EIV) is:

$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{y}}_i \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^x_i \\ \boldsymbol{\epsilon}^y_i \end{pmatrix}. \tag{1}$$

Equation (1) is only identified when multiple measurements exist for the same vector of regressors or when a subsample of observations exists in which the regressors are measured exactly (Carroll et al. (2006)). Carroll et al. (2006) discuss a standard regression setting, not a multi-input/multi-output production process. Thus, repeated measurement requires all but one of the netputs to be identical across at least two observations. Here we use the term *netputs* to describe the union of the input and output vectors. Neither of of these conditions is likely to hold for typical production data sets; therefore, we develop an alternative approach to identification.

As our starting point, we use the alternative, but equivalent, representation of the EIV model proposed by Kuosmanen and Johnson (2017):

$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{y}}_i \end{pmatrix} + e_i \begin{pmatrix} \boldsymbol{g}_i^x \\ \boldsymbol{g}_i^y \end{pmatrix}. \tag{2}$$

Clearly, the representations of Carroll et al. (2006) and Kuosmanen and Johnson (2017) are equivalent if:

$$\begin{pmatrix} \boldsymbol{\epsilon}_i^x \\ \boldsymbol{\epsilon}_i^y \end{pmatrix} = e_i \begin{pmatrix} \boldsymbol{g}_i^x \\ \boldsymbol{g}_i^y \end{pmatrix}. \tag{3}$$

We define the following normalization:

$$e_i = \sqrt{\sum_{j=1}^{d} (\epsilon_{ij}^x)^2 + \sum_{j=1}^{Q} (\epsilon_{ij}^y)^2}, \tag{4}$$

which implies:

$$\sqrt{\sum_{j=1}^{d} (g_{ij}^x)^2 + \sum_{j=1}^{Q} (g_{ij}^y)^2} = 1. \tag{5}$$

We refer to $(\boldsymbol{g}_i^x, \boldsymbol{g}_i^y)$ as the *true* noise direction and in the most general case we allow the direction to be observation specific. We note that endogeneity can exist if the error terms across variables are correlated. If the ratio of noise in different variables is common for all observations, then $(\boldsymbol{g}^x, \boldsymbol{g}^y)$ is a common direction for all observations. The perfectly correlated case is an extreme level of endogeneity. Thus, the standard EIV model with a random direction and the model with a common direction for all observations bound a variety of practical levels for this type of endogeneity.

What are the benefits of using a directional distance function versus using a standard cost or production function? The answer depends on the level of endogeneity among the variables in the production process and our assumptions about the properties of the production technology.

2.2. Production Model

Researchers use production function models, cost function models, or distance function models to characterize production technologies. Considering a general production process with multiple inputs used to produce multiple outputs, we define the production possibility set as:

$$T = \left\{ (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathbb{R}_+^{d+Q} \mid \tilde{\boldsymbol{x}} \text{ can produce } \tilde{\boldsymbol{y}} \right\}. \tag{6}$$

Following Shephard (1970), we adopt the following standard assumptions to enforce T being a production technology:

1. T is closed;
2. T is convex;

3. Free Disposability of inputs and outputs; i.e., if $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in T$ and $\left(\tilde{\boldsymbol{x}}^{\boldsymbol{k}}, -\tilde{\boldsymbol{y}}^{\boldsymbol{k}}\right) \geq (\tilde{\boldsymbol{x}}, -\tilde{\boldsymbol{y}})$, then $\left(\tilde{\boldsymbol{x}}^{\boldsymbol{k}}, \tilde{\boldsymbol{y}}^{\boldsymbol{k}}\right) \in T$.

For an alternative representation, see, for example, Frisch (1964).

Developing methods to estimate characteristics of the production technology while imposing these standard axioms was a popular and fruitful topic from the early 1950's until the early 1980's, generating such classic papers as Koopmans (1951), Shephard (1953, 1970), Afriat (1972), Charnes et al. (1978),[5] and Varian (1984). Unfortunately, these methods are deterministic in the sense that they rely on a strong assumption that the data do not contain any measurement errors, omitted variables, or other sources of random noise. Furthermore, they are often harder to implement than parametric regression. Thus, most econometricians and applied economists have chosen to use parametric models, sacrificing flexibility for ease of estimation and the inclusion of noise in the model.

Here we focus our attention on the distance function because it allows the joint production of multi-outputs using multi-inputs. The production function and cost functions can be seen as special cases of the distance function in which there is either a single output or a single input (cost) respectively. Further, motivated by our discussion of EIV models above, we consider a directional distance function which allows for measurement error in potentially all variables. We try to relax both the parametric and deterministic assumptions common in earlier approaches to modeling multi-output/multi-input technologies. We do this by building on an emerging literature that revisits the axiomatic nonparametric approach incorporating standard statistical structures including noise (Kuosmanen (2008);Kuosmanen and Johnson (2010)).

2.2.1. The Deterministic Directional Distance Function (DDF)

Luenberger (1992) and Chambers et al. (1996, 1998) introduced the directional distance function, defined for a technology T as:

$$D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) = \max\left\{\delta \in \mathbb{R} : \left(\tilde{\boldsymbol{x}} - \delta\, \boldsymbol{g}^x,\ \tilde{\boldsymbol{y}} + \delta\, \boldsymbol{g}^y\right) \in T\right\}, \tag{7}$$

where $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ are the observed input and output vectors, such that $\tilde{\boldsymbol{x}} \in \mathbb{R}^d_+$ and $\tilde{\boldsymbol{y}} \in \mathbb{R}^Q_+$ are assumed to be observed without noise and fully describe the resources used in production and the goods or services generated from production. $\boldsymbol{g}^x \in \mathbb{R}^d_+$ is the direction vector in the input space, $\boldsymbol{g}^y \in \mathbb{R}^Q_+$ is the direction vector in the output space, and $(\boldsymbol{g}^x, \boldsymbol{g}^y) \in \mathbb{R}^{d+Q}_+$ defines the direction from the point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ in which the distance function is measured.[6] δ is commonly interpreted as a measure of inefficiency by quantifying the number of bundles of size $(\boldsymbol{g}^x, \boldsymbol{g}^y)$ needed to move the observed point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ to the boundary of the technology in a deterministic setting.

[5] Data Envelopment Analysis is perhaps one of the largest success stories and has become an extremely popular method in the OR toolbox for studying efficiency.

[6] We assume $(\boldsymbol{g}^x, \boldsymbol{g}^y) \neq \mathbf{0}$; i.e., at least one of the components of either $\boldsymbol{g}^x$ or $\boldsymbol{g}^y$ is non-zero.

Chambers et al. (1998) explained how the directional distance function characterizes the technology T for a given direction vector $(\boldsymbol{g}^x, \boldsymbol{g}^y)$; specifically:

$$D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) \geq 0, \text{if and only if} \left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right) \in T. \tag{8}$$

If T satisfies the assumptions stated in Section 2.2, then the directional distance function $D_T : \mathbb{R}_+^d \times \mathbb{R}_+^Q \times \mathbb{R}_+^d \times \mathbb{R}_+^Q \to \mathbb{R}_+$ has the following properties (see Chambers et al. (1998)):

(a) $D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right)$ is upper semicontinuous in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ (jointly);

(b) $D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \lambda\, \boldsymbol{g}^x, \lambda\, \boldsymbol{g}^y\right) = (1/\lambda)\, D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right), \lambda > 0;$

(c) $\tilde{\boldsymbol{y}}' \geq \tilde{\boldsymbol{y}} \Rightarrow D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}'; \boldsymbol{g}^x, \boldsymbol{g}^y\right) \leq D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right);$

(d) $\tilde{\boldsymbol{x}}' \geq \tilde{\boldsymbol{x}} \Rightarrow D_T\left(\tilde{\boldsymbol{x}}', \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) \geq D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right);$

(e) If T is convex, then $D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right)$ is concave in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$.

An additional property of the DDF is the translation invariance:

(f) $D_T\left(\tilde{\boldsymbol{x}} - \alpha \boldsymbol{g}^x, \tilde{\boldsymbol{y}} + \alpha \boldsymbol{g}^y; \boldsymbol{g}^x, \boldsymbol{g}^y\right) = D_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) - \alpha.$

Several theoretical contributions have been made to extend the deterministic DDF, see for example Färe and Grosskopf (2010), Aparicio et al. (2017), Kapelko and Oude Lansink (2017), and Roshdi et al. (2018). The deterministic DDF has been used in several recent applications, including Baležentis and De Witte (2015), Adler and Volta (2016), and Fukuyama and Matousek (2018).

2.2.2. The Stochastic Directional Distance Function

The properties of the deterministic DDF also apply for the stochastic DDF (Färe et al. (2016)). Here we focus on estimating a stochastic DDF considering a residual which is mean zero.[7] This is represented in Figure 1.

Using the statistical model in Section 2.1 and the functional representation of technology in Section 2.2, we restate Proposition 2 in Kuosmanen and Johnson (2017) as:

Proposition 1. *If the observed data are generated according to the statistical model described in Section 2.1, then the value of the DDF in the observed data point* $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ *is equal to the realization of the random variable* ϵ_i *with mean zero, specifically*

$$D_T(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{g}^x, \boldsymbol{g}^y) = \epsilon_i \ \forall i.$$

[7] Two models are possible, 1) a mean zero residual indicating that the residual contains only noise and will pursue a productivity analysis, or 2) a composed residual with both inefficiency and noise and our direction selection analysis is for the first step of Kuosmanen and Johnson's three step procedure in which a conditional mean is estimated.

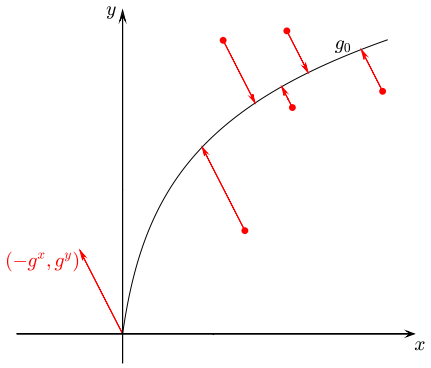

Figure 1. SDDF in mean zero case considered

In the stochastic distance function literature, the translation property, (f) above, is commonly invoked to move an arbitrarily chosen netput variable out of the distance function, yielding an equation that looks like a standard regression model; see, for example, Lovell et al. (1994) and Kuosmanen and Johnson (2017). Instead, we write the SDDF equivalently with all of the outputs on one side to emphasize that all netputs are treated symmetrically.

Under the assumption of constant returns to scale, normalizing by one of the netputs causes the noise terms to cancel for the regressors, thus eliminating the issue of endogeneity (e.g., Coelli (2000);Kuosmanen and Johnson (2017)). However, since we relax the constant returns to scale assumption, endogeneity can still be an issue.[8]

Färe et al. (2016), among others, have recognized that the selection of the direction vector affects the parameter estimates of the production function. In Appendix A.1, for the linear parametric DDF defined below, we prove that alternative directions lead to distinct parameter estimates.

3. Estimation

We now describe the estimation of the DDF under a specific parametric functional form and under nonparametric shape constrained methods.

3.1. Parametric Estimation and the DDF

Consider data composed of n observations where the inputs are defined by $\boldsymbol{x_i},\ i = 1, ..., n$ and the outputs by $\boldsymbol{y_i},\ i = 1, ..., n$. The estimator minimizes the squared residuals for a DDF with

[8] If the endogeneity is caused by correlations in the errors across variables, it can be addressed by selecting an appropriate direction for the directional distance function. This is the direction we explore in the Monte Carlo simulation section, Section 4.1.

an arbitrary prespecified direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$. For a linear production function, we formulate the estimator as:

$$\min_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{9}$$

$$\text{s.t.} \quad \boldsymbol{\gamma}' \boldsymbol{y_i} = \alpha + \boldsymbol{\beta}\, \boldsymbol{x_i} - \epsilon_i, \text{ for } i = 1, \ldots, n \tag{9a}$$

$$\boldsymbol{\beta}'\, \boldsymbol{g}^x + \boldsymbol{\gamma}'\, \boldsymbol{g}^y = 1, \tag{9b}$$

where α is the intercept, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of the marginal effects of the inputs and outputs, respectively, and the ϵ_i, $i = 1, ..., n$ are the residuals.

Equation (9b) enforces the translation property described in Chambers et al. (1998); i.e., scaling the netput vector by δ in the direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$ causes the distance function to decrease by δ. The combination of Equation (9a) and Equation (9b) ensures that the residual is computed along the direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$.

3.2. The CNLS-d Estimator

Convex Nonparametric Least Squares (CNLS) is a non-parametric estimator that imposes the axiomatic properties, such as monotonicity and concavity, on the production technology. The estimator CNLS-d is the directional distance function generalization of CNLS (Hildreth (1954); Kuosmanen (2008)). While CNLS allows for just a single output, CNLS-d permits multiple outputs. In CNLS the direction along which residuals are computed is specified *a priori* and is typically measured in terms of the unique output, y. This corresponds to the assumption that noise is only present in y and that all other variables, $\tilde{\boldsymbol{x}}$, do not contain noise. CNLS-d allows the residual to be measured in an arbitrary prespecified direction. If all components of the direction vector are non-zero, this corresponds to an assumption that noise is present in all inputs.

Using the same input-output data defined in 2.1, the CNLS-d estimator is given by:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{10}$$

$$\text{s.t.} \quad \boldsymbol{\gamma_i'}\, \boldsymbol{y_i} = \alpha_i + \boldsymbol{\beta_i'}\, \boldsymbol{x_i} - \epsilon_i, \text{ for } i = 1, \ldots, n \tag{10a}$$

$$\alpha_i + \boldsymbol{\beta_i'}\, \boldsymbol{x_i} - \boldsymbol{\gamma_i'}\, \boldsymbol{y_i} \leq \alpha_j + \boldsymbol{\beta_j'}\, \boldsymbol{x_i} - \boldsymbol{\gamma_j'}\, \boldsymbol{y_i}, \text{ for } i, j = 1, \ldots, n,\ i \neq j \tag{10b}$$

$$\boldsymbol{\beta_i} \geq 0, \text{ for } i = 1, \ldots, n \tag{10c}$$

$$\boldsymbol{\beta_i'}\, \boldsymbol{g}^x + \boldsymbol{\gamma_i'}\, \boldsymbol{g}^y = 1, \text{ for } i = 1, \ldots, n \tag{10d}$$

$$\boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1, \ldots, n, \tag{10e}$$

where α_i, $i = 1, ..., n$ is the vector of the intercept terms, $\boldsymbol{\beta_i}$, $i = 1, .., n$ and $\boldsymbol{\gamma_i}$, $i = 1, .., n$ are the

matrices of the marginal effects of the inputs and the outputs, respectively, and ϵ_i, $i = 1, ..., n$ is the vector of the residuals (Kuosmanen and Johnson, 2017).

Equation (10b), which corresponds to the Afriat inequalities, imposes concavity. Given Equation (10b), Equation (10c) imposes the monotonicity of the estimated frontier relative to the inputs. Equation (10d) enforces the translation property described in Chambers et al. (1998) and has the same interpretation as Equation (9b). Similar to Equation (10c), the combination of Equation (10b) and Equation (10e) imposes the monotonicity of the DDF relative to the outputs. In Equation (10), we specify the CNLS-d estimator with a single common direction, $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$.[9]

4. Measuring MSE under Alternative Directions

4.1. Illustrative Example

Data Generation Process. For our illustrative example, we use a simple linear cost function and a directional distance linear parametric estimator. We consider two noise generation processes: a random noise direction and a fixed noise direction (the latter corresponds to a case of severe endogeneity). Here we discuss the random noise direction case, but direct the reader to Appendix B for a discussion of the fixed noise direction case.

For our example, we consider the following single output cost function Data Generation Process (DGP) where the observations $(y_i, c_i), i = 1, \ldots, n$ are generated as outlined in Figure 2:

[9] Alternatively, some researchers may be interested in using observation specific directions or perhaps group specific directions. In Appendix A.3, we derive the conditions under which multiple directions can be used in CNLS-d while still maintaining the axiomatic property of global convexity of the production technology. Consider two groups each with their own direction used in the directional distance function. Essentially, the convexity constraint holds as long as the noise is orthogonal to the difference of the two directions used in the estimation. A simple example of this situation is all the noise being in one dimension and the difference between the two directions for this dimension is zero. However, this condition is restrictive when noise is potentially present in all variables. Thus, specifying multiple directions in CNLS-d while maintaining the axiomatic properties of the estimator, specifically, the convexity of the production possibility set, is still an open research question.

1. Output, $\tilde{y}_i$, is drawn from the continuous uniform distribution $U[0,1]$.

2. Cost is calculated as $\tilde{c}_i = \beta_0\ \tilde{y}_i$, where $\beta_0 = 1$.

3. The noise terms, $\epsilon_{y_i}, \epsilon_{c_i}$, are constructed as follows:

   (a) ϵ_0 is calculated as:

   $$\epsilon_0 = \frac{1}{2}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2} + \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{c}_i - \bar{c}\right)^2}\right], \tag{11}$$

   where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i$ and $\bar{c} = \frac{1}{n}\sum_{i=1}^{n}\tilde{c}_i$ are the means of the output and cost without noise, respectively.

   (b) The scalar length of the noise is rescaled by the vector, $v_{q\epsilon_i}$, in each dimension. These scaling factors are calculated as $v_{q\epsilon_i} = \frac{v^*_{q\epsilon_i}}{L\left(\boldsymbol{v}^*_{\boldsymbol{\epsilon_i}}\right)}, q = \{1,2\}$ where $v^*_{q\epsilon_i}$ are drawn from a continuous uniform distribution $U[-1,1]$ and $L\left(\boldsymbol{v}^*_{\boldsymbol{\epsilon_i}}\right) = \sqrt{\sum_{q=1}^{2}\left(v^*_{q\epsilon_i}\right)^2}$.

   (c) $\left(\epsilon_{y_i}, \epsilon_{c_i}\right) = l_{\epsilon_i}\,\boldsymbol{v}_{\boldsymbol{\epsilon_i}}, i = 1,\ldots,n$, where l_{ϵ_i} is a scalar length drawn from the normal distribution, $N\left(0, \lambda\,\epsilon_0\right)$, where λ is prespecified initial value for the standard deviation and $\boldsymbol{v}_{\boldsymbol{\epsilon_i}} = [v_{1\epsilon_i}, v_{2\epsilon_i}]$ is a normalized direction vector.

4. The observations with noise are simply obtained by adding the noise term:

   $$\begin{pmatrix} y_i \\ c_i \end{pmatrix} = \begin{pmatrix} \tilde{y}_i \\ \tilde{c}_i \end{pmatrix} + \begin{pmatrix} \epsilon_{y_i} \\ \epsilon_{c_i} \end{pmatrix}, i = 1,\ldots,n. \tag{12}$$

Figure 2. Linear function data generation process with random noise directions

Figure 3 show the results of the two data generating processes, the case in which the direction of the noise is random and the case in which the direction of the noise of fixed, respectively.

Evaluating the Parametric Estimator's Performance. We use two criteria to assess the performance of the parametric estimator: 1) Mean Squared Error (MSE) comparing the true function to the estimated function, and 2) MSE comparing the estimated function to a testing dataset. While we can calculate both metrics for our Monte Carlo simulations, only the second metric can be used with our application data below.

To calculate deviations, we use the MSE direction $\left(g^y_{MSE}, g^c_{MSE}\right)$. For any particular point of the testing set, $\left(y_{ts_i}, c_{ts_i}\right), i = 1,\ldots,n$, we determine the estimates, $\left(\hat{y}_{ts_i}, \hat{c}_{ts_i}\right), i = 1,\ldots,n$, defined as the intersection of the estimated function characterized by the coefficients $\left(\hat{\alpha}, \hat{\beta}\right)$ and the line passing through $\left(y_{ts_i}, c_{ts_i}\right), i = 1,\ldots,n$, and direction vector $\left(g^y_{MSE}, g^c_{MSE}\right)$. We evaluate

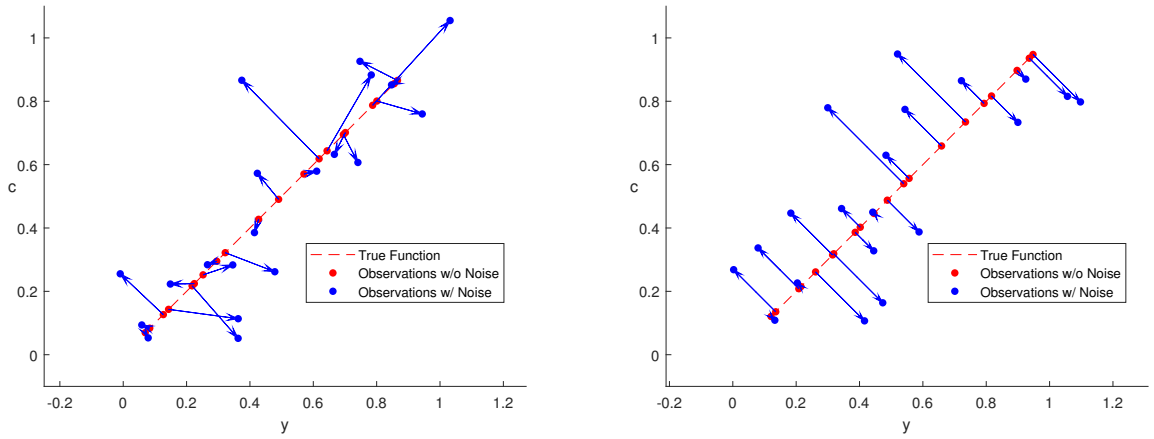

Figure 3. Linear Case with Random Noise Direction (left), Linear Case with Fixed Noise Direction (right)

the value of the MSE as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left((\hat{y}_{ts_i} - y_{ts_i})^2 + (\hat{c}_{ts_i} - c_{ts_i})^2\right). \tag{13}$$

We use the same method when comparing the true function to the estimated function and the estimated function to our testing dataset. To compare the true function to the estimated function, we use the Linear Function Data Generation Process, the left figure of Figure 3 steps 1 and 2, to construct our testing dataset, and to evaluate the estimated function without knowing the true function using the full Linear Function Data Generation Process. Figure 4 show the MSE computations.

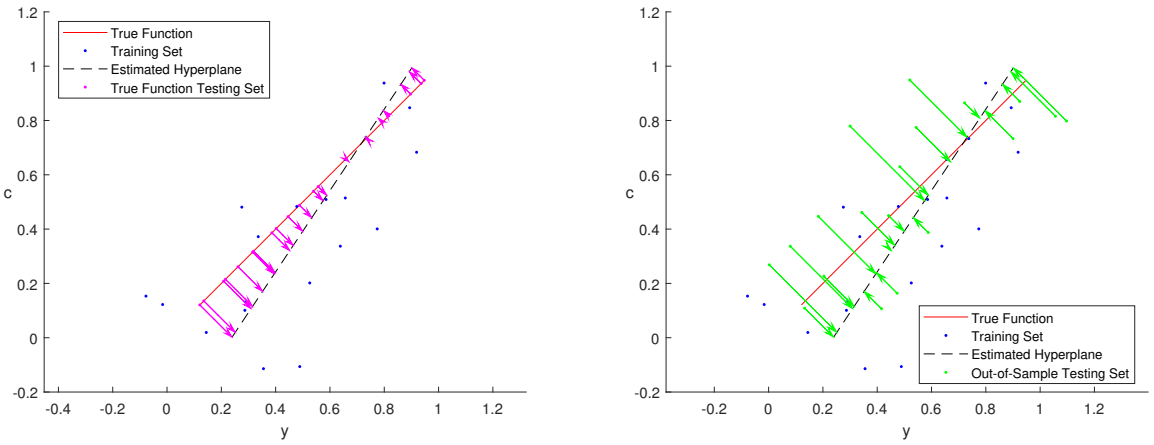

Figure 4. MSE calculated relative to the True Function in the MSE direction $\pi/4$ (left), MSE calculated using a testing data set in the MSE direction $\pi/4$ (right)

Additional Information Describing the Simulations. We apply the DGP described above to generate a training set, $(y_{tr_i}, c_{tr_i}), i = 1, \ldots, n_{tr}$, and a testing set $(y_{ts_i}, c_{ts_i}), i = 1, \ldots, n_{ts}$, in which noise is introduced to the observations in random directions. We set the noise scaling coefficient to $\lambda = 0.6$ and the number of observations to $n_{tr} = n_{ts} = 100$. We run 100 repetitions of the simulation for each experiment on a computer with a processor Intel Core i7 CPU 860 2.80 GHz and 8 GB RAM. We use the quadratic solver on MATLAB 2017a.

For the estimator, we define the direction vector used in the parametric DDF as a function of an angular variable θ, which allows us to investigate alternative directions. Specifically, the direction vector used in the DDF is $(g^y, g^c) = (\cos(\theta_t), \sin(\theta_t))$. We examine the set of directions corresponding to the angles $\theta_t \in \{0,\ \pi/8,\ \pi/4,\ 3\pi/8,\ \pi/2\}$.

Results: Random Noise Directions. Table 1 and Table 2 show results corresponding to the two performance criteria introduced above, the MSE relative to the true function and the MSE relative to a testing dataset, respectively. Table 1 shows that the direction corresponding to the angle $\pi/4$, $(g^y = 0.707, g^c = 0.707)$, produces the smallest values of MSE (shown in bold in the table) regardless of the direction used for the MSE computation. Thus, for noise introduced in random directions and given this particular DGP, we are able to fit an optimal direction for estimation. However, the estimator's quality diminishes if we select the extreme directions corresponding to the angles 0 and $\pi/2$. Table 2 reports that the direction corresponding to the smallest MSE value (shown in bold) is always the one matching the direction used in the MSE computation. In applications, using a testing set is necessary because the true function is unknown. Table 2 shows the benefits of matching the direction of MSE evaluation direction outweigh the benefits of selecting a direction based on the properties of the function being estimated.

Table 1. Average MSE over 100 simulations for the Linear Estimator compared to the true function with a DGP using random noise directions

	Avg MSE: Comparison to the True Function DDF Angle θ_t				
MSE Dir Angle θ_{MSE}	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	2.09	0.75	**0.56**	1.16	3.68
$\pi/8$	1.36	0.46	**0.32**	0.63	1.89
$\pi/4$	1.25	0.41	**0.28**	0.51	1.48
$3\pi/8$	1.59	0.50	**0.32**	0.57	1.60
$\pi/2$	3.06	0.91	**0.55**	0.92	2.44

Note: Displayed are measured values multiplied by 10^3.

For the out-of-sample testing set, the direction that provides the smallest MSE value is the direction used for the MSE computation. Because the functional estimate is optimized for the direction specified in the SDDF, it is perhaps expected that using the same direction that will be

Table 2. Average MSE over 100 simulations for the Linear Estimator compared to an out-of-sample testing set with a DGP using random noise directions

	Avg MSE: Comparison to Out-of-Sample DDF Angle θ_t				
MSE Dir Angle θ_{MSE}	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	**28.28**	29.43	31.29	34.23	40.67
$\pi/8$	18.03	**17.79**	18.19	19.09	21.32
$\pi/4$	16.38	15.55	**15.45**	15.77	16.90
$3\pi/8$	20.50	18.67	18.04	**17.90**	18.46
$\pi/2$	38.63	33.07	30.68	29.29	**28.70**

Note: Displayed are measured values multiplied by 10^3.

used in the MSE evaluation would produce a relatively low MSE compared to other directions. However, when the functional estimate is compared to the true function, the MSE values are around ten times smaller than the out-of-sample testing case. In out-of-sample testing the presence of noise in the observations causes a deviation regardless of the quality of the estimator or the number of observations. The DDF direction corresponding to the smallest MSE is the direction orthogonal to the true function (i.e., $\pi/4$ for our DGP). This direction provides the shortest distance from the observations to the true function. We conclude that, in this experiment, it is preferable to select a direction orthogonal to the true function (see Section 5 for further experiments).

From the fixed noise direction experiments (see Appendix B), we observe that using a direction for the estimator that matches the direction used for the noise generation significantly reduces the MSE values compared to the true function. From this, we infer that when endogeneity is severe, using a direction that matches the characteristics of this endogeneity significantly improves the fit of the estimator; i.e., the MSE is 50% smaller for the matching direction than for the second best direction in 70% of the cases (see Section 5 for the details).

Finally, we need to solve the problem of evaluating alternative directions when the true function is unknown so that we can evaluate alternative directions in the application data. Below, we describe our proposed alternative measure of fit.

4.2. Radial MSE Measure

MSE is typically measured by the average sum of squared errors in the dimension of a single variable, such as cost or output. As explained in Section 4.1, when we compare out-of-sample performance, we find that the best direction to use in estimating a SDDF is the direction used for MSE evaluation regardless of the direction of noise in the DGP or any other characteristics of the DGP. To avoid this relationship between the direction of estimation and the direction of evaluation, we propose a **radial MSE measure**.

We begin by normalizing the data to a unit cube and consider a case of Q outputs and n observations, where the original observations are:

$$(y_{i1}, \ldots, y_{iQ}, c_i), \ i = 1, \ldots, n.$$

The normalized observations are:

$$\breve{y}_{ij} = \frac{y_{ij} - \min_k y_{kj}}{\max_k y_{kj} - \min_k y_{kj}}, \ j = 1, \ldots, Q, \ i = 1, \ldots, n, \tag{14}$$

$$\breve{c}_i = \frac{y_i - \min_k c_k}{\max_k c_k - \min_k c_k}, \ i = 1, \ldots, n. \tag{15}$$

Our radial MSE measure is the distance from the testing set observation to the estimated function measured along a ray from the testing set observations to the **center** C. Having normalized the data, the center for the radial measure is $C = \left[\overbrace{0, \ldots, 0}^{Q}, 1\right]$.

The radial MSE measure is the average of the distance from each testing set observation to the estimated function measured radially. Figure 5 illustrates this measure. For a convex function, a radial measure reduces the bias in the measure for extreme values in the domain.

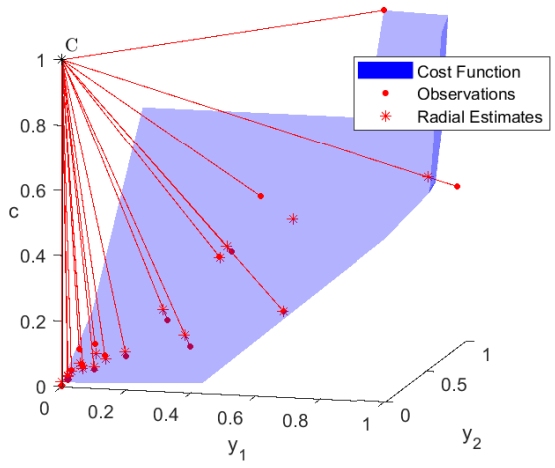

Figure 5. A Radial MSE Measure on a Cost Function with Two Outputs

5. Monte Carlo Simulations

We next examine how different DGPs affect the optimal direction for the DDF estimator based on a set of Monte Carlo simulations. We consider both random noise directions for each observation and a fixed noise direction representing a high endogeneity case. We consider the effects of the different variance levels for the noise and changes in the underlying distribution of the production data. Using the simplest case of two outputs and a fixed cost level for all observed units allows us to isolate the effects of the characteristics of the data and of the function separately.

5.1. CNLS-d Formulation for Cost Isoquant Estimation

Before describing our experiments, we first outline the CNLS-d for estimating the iso-cost level set. It is based on the following optimization problem:

$$\min_{\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{16}$$

$$\text{s.t.} \quad -\epsilon_j + \epsilon_i - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \leq 0, \text{ for } i,j = 1,\ldots,n,\ i \neq j \tag{16a}$$

$$\boldsymbol{\gamma_i'}\, \boldsymbol{g}^y = 1, \text{ for } i = 1,\ldots,n \tag{16b}$$

$$\boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1,\ldots,n. \tag{16c}$$

We can recover the fitted values, $\hat{y}_i$, and the coefficient, $\alpha_i,\ i = 1,\ldots,n$, using:

$$\boldsymbol{\hat{y}_i} = \boldsymbol{y_i} - \epsilon_i\, \boldsymbol{g}^y, \text{ for } i = 1,\ldots,n \tag{17}$$

$$\alpha_i = \boldsymbol{\gamma_i'}\, \boldsymbol{y_i} + \epsilon_i, \text{ for } i = 1,\ldots,n. \tag{18}$$

5.2. Experiments

We conducted four experiments to investigate the optimal direction for the DDF estimator.

Experiment 1 - Base case: A two output circular isoquant with uniformly distributed angle parameters and random noise direction.

For the base case, we consider a fixed input level and approximate a two output isoquant; i.e., $Q = 2$. Indexing the outputs by q and observations by i, we generate the output variables as:

$$y_{qi} = \tilde{y}_{qi} + \epsilon_{qi},\ q = 1,\ldots,Q,\ i = 1,\ldots,n, \tag{19}$$

where $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ is the observation on the isoquant and $\boldsymbol{\epsilon_i}$ is the noise. We generate the output levels $\tilde{y}_{qi},\ q = 1,\ldots,Q\ , i = 1,\ldots,n$ as:

$$\tilde{y}_{1i} = \cos(\theta_i),\ i = 1, \ldots, n \tag{20}$$

$$\tilde{y}_{2i} = \sin(\theta_i),\ i = 1, \ldots, n, \tag{21}$$

where $\theta_i,\ i = 1, \ldots, n$, is drawn randomly from a continuous uniform distribution, $U\left[0, \frac{\pi}{2}\right]$. The noise terms, $\epsilon_{qi},\ q = 1, \ldots, Q,\ i = 1, \ldots, n$, have the following expressions:

$$\epsilon_{1i} = l\ \cos(\theta_{\epsilon_i}),\ i = 1, \ldots, n \tag{22}$$

$$\epsilon_{2i} = l\ \sin(\theta_{\epsilon_i}),\ i = 1, \ldots, n, \tag{23}$$

where the length l is drawn from the normal distribution $N(0, \lambda)$, the angle θ_{ϵ_i} is observation specific and characterizes the noise direction for each observation, and θ_{ϵ_i} is drawn from a continuous uniform distribution $U\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The values considered for the directions in CNLS-d estimator are $0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}$. The standard deviation of the normal distribution is $\lambda = 0.1$. We perform the experiment 100 times for each parameter setting.

Table 3 reports the radial MSE values from a testing set of n observations lying on the true function.

Table 3. Experiment 1: Values of the radial MSE relative to the true function. The angle used in CNLS-d estimator varies and the noise direction is randomly selected. In the DGP, the standard deviation of the noise distribution, λ, is 0.1.

	CNLS-d Direction Angle				
	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
Average MSE across simulations	13.90	4.65	**3.32**	4.49	13.93

Note: Displayed are measured values multiplied by 10^4.

As shown in Table 3, the angle corresponding to the smallest MSE (shown in bold) is the one that gives an orthogonal direction to the center of the true function, $\frac{\pi}{4}$, and that the MSE values differ significantly, increasing at similar rates as the direction angle deviates from $\frac{\pi}{4}$ in either direction.

Experiment 2 - The base case with fixed noise directions.

In this experiment, θ_{ϵ_i}, which characterizes the noise direction for each observation, is constant for all observations, θ_ϵ. The values used for θ_ϵ and the directions in CNLS-d estimator are the same, $0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}$. The standard deviation of the normal distribution is again $\lambda = 0.1$. We perform the experiment 100 times for each parameter settings. Table 4 reports the results.

Each row in the Table 4 corresponds to a different noise direction in DGP. The bold numbers identify the directions in CNLS-d estimator that obtains the smallest MSE for each noise direction.

We confirm our previous insight, from the parametric estimator and fixed noise direction case described in Appendix B, that the bold values appearing on the diagonal (from the upper-left to the lower-right of Table 4) correspond to the directions used in CNLS-d. This result indicates that selecting the direction in the SDDF that matches the underlying noise direction in the DGP results in improved functional estimates.

Table 4. Experiment 2: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d estimator direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.1.

	CNLS-d Direction Angle				
Noise Direction Angle	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	**2.69**	3.03	4.49	8.86	25.47
$\pi/8$	7.49	**3.44**	4.00	8.07	28.83
$\pi/4$	20.28	5.79	**4.30**	5.80	19.06
$3\pi/8$	25.58	7.80	4.18	**3.51**	6.84
$\pi/2$	25.90	9.09	4.73	3.10	**2.57**

Note: Displayed are measured values multiplied by 10^4.

Experiment 3. Base case with fixed noise direction and different noise levels.
In Experiment 3, we vary the noise term by changing the λ coefficient from $\lambda = 0.1$ to a value half as large, $\lambda = 0.05$, and a value twice as large, $\lambda = 0.2$. Table 5 reports the results for $\lambda = 0.05$:

Table 5. Experiment 3–Less Noise: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.05.

	CNLS-d Direction Angle				
Noise Direction Angle	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	0.92	**0.82**	0.96	1.53	5.12
$\pi/8$	1.83	**1.09**	**1.09**	1.47	5.45
$\pi/4$	3.70	1.41	**1.29**	1.43	3.93
$3\pi/8$	5.75	1.68	1.27	**1.18**	1.86
$\pi/2$	4.61	1.40	0.95	**0.79**	0.90

Note: Displayed are measured values multiplied by 10^4.

In Table 5 (Experiment 3, with $\lambda = 0.05$), we do not observe the same diagonal pattern observed in Experiment 2, and the best direction for CNLS-d estimator does not match the direction selected for the noise. This leads us to hypothesize that when the noise level is small, data characteristics, such as the distribution of the regressors or the shape of the function, affect the estimation whereas when the noise level is large, regressors' relative variability becomes a more dominant factor in determining the best direction for the CNSL-d estimator.

The results of Experiment 3 with $\lambda = 0.2$ are reported in the appendix, section Appendix B, Table B.15 (Experiment 3 with $\lambda = 0.2$). The results are consistent with the results of Experiment 2, specifically, the best direction always coincides with the noise direction selected.

Experiment 4: Base case with different distributions for the initial observations on the true function. In Experiment 4, we seek to understand how changing the DGP for the angle, $\theta_i,\ i = 1, \ldots, n$, affects the optimal direction. We consider the three normal distributions with different parameters: $N\left[\frac{\pi}{8}, \frac{\pi}{16}\right]$, $N\left[\frac{\pi}{4}, \frac{\pi}{16}\right]$ and $N\left[\frac{3\pi}{8}, \frac{\pi}{16}\right]$. We truncate the tails of the distribution so that the generated angles fall in the range $[0, \pi/2]$. Noise is specified as in Experiment 1. Table 6 reports the results of this experiment.

Table 6. Experiment 4: Values of radial MSE relative to the true function varying the CNLS-d direction and the mean of the normal distribution used in the DGP.

Mean of the	CNLS-d Direction angle				
Normal Distribution ($\bar{\theta}$)	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
$\pi/8$	3.19	**2.21**	3.89	10.28	46.47
$\pi/4$	8.44	2.92	**1.98**	3.17	9.00
$3\pi/8$	45.64	10.25	4.02	**2.43**	3.07

Note: Displayed are measured values multiplied by 10^4.

In Table 6, we observe that selecting a direction in the SDDF to match $\bar{\theta}$, the mean of the distribution for the angle variable used in the DGP, corresponds to the smallest MSE value. This result suggests that the estimator's performance improves when we select a direction that points to the "center" of the data.

6. Proposed Approach to Direction Selection

Based on Monte-Carlo simulations, we have found that the optimal direction depends on the shape of the function and the distribution of the observed data. This of itself is not surprising. However, by assuming a unimodal distribution for the data generation process, a direction that aims towards the "center" of the data and is perpendicular to the true function at that point tends to outperform other directions. To apply this finding for a dataset with Q outputs and n observations, $(y_{i1}, \ldots, y_{iQ}, c_i),\ i = 1, \ldots, n$, we suggest selecting the direction for the DDF as follows:

1. Normalize the data:

$$\breve{y}_{ij} = \frac{y_{ij} - \min_k y_{kj}}{\max_k y_{kj} - \min_k y_{kj}},\ j = 1, \ldots, Q,\ i = 1, \ldots, n \tag{24}$$

$$\breve{c}_i = \frac{y_i - \min_k c_k}{\max_k c_k - \min_k c_k},\ i = 1, \ldots, n \tag{25}$$

2. Select the direction:

$$\begin{bmatrix} g_{y1} \\ \vdots \\ g_{yQ} \\ g_c \end{bmatrix} = \begin{bmatrix} \text{median}\left(\breve{y}_{i1}\right) \\ \vdots \\ \text{median}\left(\breve{y}_{iQ}\right) \\ 1 - \text{median}\left(\breve{c}_i\right). \end{bmatrix} \tag{26}$$

This provides a method for direction selection that can be used in applications when the true direction is unknown. We test the proposed method by estimating a cost function for a US hospital dataset.

7. Cost Function Estimation of the US Hospital Sector

We analyze the cost variation across US hospitals using a conditional mean estimate of the cost function. We estimate a multi-output cost function for the US hospital sector by implementing our data-driven method for selecting the direction vector for the DDF. We report most productive scale size and marginal cost estimates.

7.1. Description of the Dataset

We obtain cost data from the American Hospital Association's (AHA) Annual Survey Databases from 2007 to 2009. The costs reported include payroll, employee benefits, depreciation, interest, and supply expenses. We obtain hospital output data from the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) core file that captures data annually for all discharges for a 20% sample of US community hospitals. The hospital sample changes every year. For each patient discharged, all procedures received are recorded as International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9-CM) codes. The typical hospital in the US relies on these detailed codes to quantify the medical services it provides and summarize the resources consumed via cost measures (Zuckerman et al. (1994)). We map the codes to four categories of procedures, specifically the procedure categories are "Minor Diagnostic," "Minor Therapeutic," "Major Diagnostic," and "Major Therapeutic" which are standard output categories in the literature (Pope and Johnson (2013)). The number of procedures is each category are summed for each hospital by year to construct the output variables. The total number of hospitals sampled is

around 1,000 per year from 2007 to 2009.[10] However, mapping between the two databases is only possible for approximately 50% of the hospitals in the HCUP data, resulting in approximately 450 to 525 observations available each year.

Table 7. Summary Statistics of the hospital dataset

	2007 (523 observations)				
	Cost ($)	MajDiag	MajTher	MinDiag	MinTher
Mean	146M	162	4083	3499	7299
Skewness	3.51	2.89	2.63	5.19	3.28
25-percentile	24M	9	277	108	512
50-percentile	72M	73	1688	938	3108
75-percentile	182M	207	5443	4082	9628
	2008 (511 observations)				
	Cost ($)	MajDiag	MajTher	MinDiag	MinTher
Mean	163M	175	4433	3688	7657
Skewness	4.19	3.80	2.97	4.87	2.82
25-percentile	28M	10	325	120	545
50-percentile	83M	76	1809	1013	3350
75-percentile	189M	246	5984	4569	10781
	2009 (458 observations)				
	Cost ($)	MajDiag	MajTher	MinDiag	MinTher
Mean	175M	161	4471	3615	7905
Skewness	3.39	3.78	2.43	4.68	2.41
25-percentile	31M	12	420	148	713
50-percentile	91M	69	1737	1136	3458
75-percentile	220M	230	6402	4694	10989

7.2. Pre-Analysis of the Dataset

7.2.1. Testing the Relevance of the Regressors

We begin by testing the statistical significance of our four output variables, $\boldsymbol{y} = (y_1, y_2, y_3, y_4)$, for predicting cost. The null hypothesis stated for the qth output is:

$$H_0 : P\left[E\left(c \,|\, \boldsymbol{y} - \{y_q\}\right) = E\left(c \,|\, \boldsymbol{y}\right)\right] = 1$$

[10]The NIS survey is a stratified systematic random sample. The strata criteria are urban or rural location, teaching status, ownership, and bed size. This stratification ensures a more representative sample of discharges than a simple random sample would yield. For details see `https://www.hcup-us.ahrq.gov/tech_assist/sampledesign/508_compliance/508course.htm#{463754B8-A305-47E3-B7EE-A43953AA9478}`.

against:[11]

$$H_1 : P\left[E\left(c \,|\, \boldsymbol{y} - \{y_q\}\right) = E\left(c \,|\, \boldsymbol{y}\right)\right] < 1.$$

We implement the test with a Local Constant Least Squares (LCLS) estimator described in Henderson and Parmeter (2015), calculating bandwidths using least-squares cross-validation. We use 399 wild bootstraps. We found that all output variables were highly statistically significant in all years.

7.3. Results

CNLS-d and Different Directions. We analyze each year of data as a separate cross-section because, as noted above, the HCUP does not track the same set of hospitals across years. To illuminate the direction's effect on the functional estimates, we graph "Cost" as a function of "Major Diagnostic Procedures" and "Major Therapeutic Procedures" holding "Minor Diagnostic Procedures" and "Minor Therapeutic Procedures" constant at their median values. Figure 6 illustrates the estimates for three different directions, one with only a cost component, one with only a component in Major Therapeutic Procedures, and one that comes from our median approach. Visual inspection indicates that the estimates with different directions produce significantly different estimates, highlighting the importance of considering the question of direction selection.

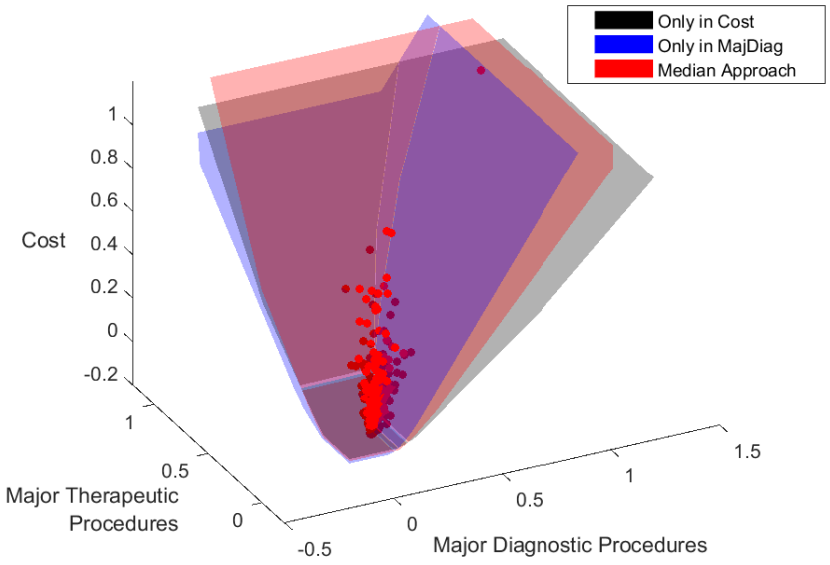

Figure 6. US Hospital Cost Function Estimates for Three Directions

[11] Where the notation $\boldsymbol{y} - \{y_q\}$ implies the vector $\boldsymbol{y}$ excluding the qth component.

We compare the estimator's performance when using different directions. Table 8 reports the MSE for three sample directions in each year. We define our direction vector as $(g_{y1}, g_{y2}, g_{y3}, g_{y4}, g_c)$.[12]

Table 8. Results of the radial MSE values for different directions by year

Direction	Year		
$(g_{y1}, g_{y2}, g_{y3}, g_{y4}, g_c)$	2007	2008	2009
(0.45, 0.45, 0.45, 0.45, 0.45)	2.10	**1.30**	1.50
(0.35, 0.35, 0.35, 0.35, 0.71)	2.15	1.65	**1.29**
Median Direction	**1.79**	1.55	1.34

Note: Displayed are the measured values multiplied by 10^3

We pick two directions, one with equal components in all dimensions, and a second direction that has a cost component that is double the value of the output components. The median vector is $(0.014, 0.041, 0.033, 0.038, 0.998)$, which is very close to the cost-only direction. The MSE varies by 15-30% over the different directions. We observe that there is no clear dominate direction; however, the median direction performs reasonably well in all cases. We conclude that as long as a direction with non-zero components for all variables that could contain noise is selected, then the precise direction selected is not critical to obtaining improved estimation results.

Comparison with other estimators. We compare three methods to estimate a cost function: 1) a quadratic functional form (without the cross-product terms), Färe et al. (2010); 2) CNLS-d with the direction selection method proposed in Section 6; and 3) Lower bound estimate calculated using a Local Linear Kernel regression with a Gaussian kernel and Leave One-Out Cross-validation for bandwidth selection, Li and Racine (2007).[13] We select these estimators because a quadratic functional form to model production has been used in recent productivity and efficiency analysis of healthcare. See, for example, Ferrier et al. (2018). The local linear kernel is selected because it is an extremely flexible nonparametric estimator and provides a lower bound for the performance of a functional estimate. However, note the local linear Kernel does not have the standard properties of a cost function, specificaly cost is monotonic in cost and marginal costs are increasing as output increases.

We will use the criteria of K-fold average MSE with $k = 5$ to compare the approaches. This means we split the data equally into 5 parts. We use 4 of the 5 parts for estimation (training) and evaluate the performance of the estimator on the 5th part (testing). We do this for all 5 parts and average the results. The values presented correspond to the average across folds. Table 9 reports the results.

[12] We focus on types of directions found to be competitive in our Monte Carlo simulations.

[13] For CNLS-d, we select a value for an upper bound through a tuning process, Ubound = 0.5, and impose the upper bound on the slope coefficients estimated (Lim, 2014).

Table 9. US Hospital K-fold Average MSE in Cost to the Cost Function Estimates for the Three Functional Specifications by Year

Year	Quadratic Regression	CNLS-d (Median Direction)	Lower Bound Estimator
2007	3.43	2.44	2.35
2008	2.76	1.93	1.48
2009	2.43	1.80	1.53

Note: The MSE values displayed are the measured values multiplied by 10^3

Overall CNLS-d performs well and is close to the lower bound in terms of fitting performance while imposing standard axioms of a cost function. Like most production data, the hospital data is very noisy. The shape restrictions imposed in CNLS-d improves the interpretability. The CNLS-d estimator outperform the parametric approach, indicating the general benefits of nonparametric estimators.

Description of Functional Estimates - MPSS and Marginal Costs. We report the Most Productive Scale Size (MPSS) and the Marginal Costs for the a quadratic parametric estimator, the CNLS-d estimator with our proposed direction selection method and an alternative. For the MPSS, we present the cost levels obtained for different ratios of *Major Diagnotic procedures* (MajDiag) and *Major Therapeutic procedures* (MajTher), with the minor diagnostics and therapeutics held constant at their median levels. For the Marginal Costs, we present the values for different percentiles of the MajDiag and MajTher, with the minor diagnostics and therapeutics held constant at their median levels. A more exhaustive comparison across all outputs is presented in Appendix C. MPSS results are presented in Table 10 and the values for CNLS-d (Median Direction) are illustrated in Figure 7. We observe small variations across both years and estimators. The differences across years are in part due to the sample changing across years. Most hospitals are small and operate close to the MPSS.[14] However, there are several large hospitals that are operating significantly above MPSS. Hospitals might choose to operate at larger scales due to synergies of joint production and network effects, meaning providing a large array of services allows hospitals to increase their production levels by providing one location where a consumer can fulfill multiple healthcare needs.

CNLS-d is the most flexible estimator and allows MPSS values to fluctuate significantly across percentiles. CNLS-d does not smooth variation, rather it minimizes the distance from each observations to the shape constrained estimator. Even though the LL Kernel bandwidths are selected via cross-validation, relatively large values are selected due to the relatively noisy data and the highly skewed output distributions. These large bandwidths and the parametric nature of the Quadratic

[14]Recall the data has been rescaled so that all variables including cost are on the scale of 0 to 1.

function make these two estimators relatively less flexible compared to CNLS-d. A feature that is only captured by CNLS-d is that in all years hospitals that specialize in either major diagnostics or major therapeutics maximize productivity at a larger scale of operation.

Table 10. Most Productive Scale Size measured in normalized cost conditional on Major Diagnostic procedures (MajDiag) and Major Therapeutic procedures (MajTher) , Minor Diagnostic procedures (MinDiag) and Minor Therapeutic procedures (MinTher) held constant at the 50th percentile

Ratio	Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MajTher/MajDiag	2007	2008	2009	2007	2008	2009	2007	2008	2009
20%	0.73	1.69	1.46	0.96	0.46	1.02	0.88	0.45	1.01
30%	0.87	1.22	1.28	0.46	0.44	1.01	0.83	0.42	0.49
40%	1.00	1.08	1.19	0.43	0.41	0.48	0.41	0.39	0.47
50%	1.13	0.81	1.34	0.43	0.39	0.45	0.42	0.38	0.44
60%	1.25	0.87	1.08	0.46	0.39	0.43	0.46	0.38	0.44
70%	0.89	0.91	1.14	0.98	0.42	0.45	0.95	0.41	0.44
80%	0.93	0.93	1.17	1.01	0.44	0.96	0.99	0.43	0.45

Note: The values displayed are the measured values multiplied by 10

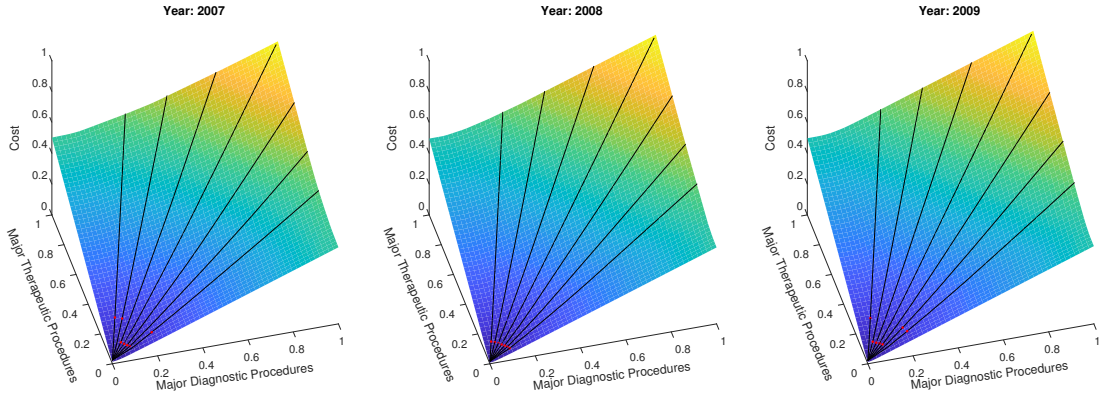

Figure 7. Most Productive Scale Size (in red) on the estimated function by "CNLS-d (Med)", CNLS-d using the median approach for the direction, for different ratios of Major Therapeutic Procedures over Major Diagnostic Procedures

The marginal cost results for Major Diagnostic Procedures are presented in Table 11 and Figure 8 (left) and the marginal cost results for Major Therapeutic Procedures in Table 12 and Figure 8 (right). Again, as in Table 10, reporting MPSS levels, CNLS-d is more flexible and its marginal cost estimates vary significantly across percentiles. The CNLS-d with difference directions provides very similar marginal costs estimates. However, the CNLS-d estimates differ significantly from the marginal cost estimates from the parametric estimator. It seems that the direction choice is not as critical as long as the direction contains a non-zero component in all the netputs that contribute to

Table 11. Marginal Cost of Major Diagnostic Procedures

Percentile		Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	0.08	1.26	0.92	0.02	0.03	0.03	0.05	0.07	0.06
25	50	0.08	1.26	0.92	0.09	0.02	0.03	0.02	0.03	0.06
25	75	0.08	1.26	0.92	0.02	0.04	0.04	0.03	0.01	0.06
50	25	0.24	1.25	0.92	2.09	2.86	2.30	2.56	1.51	2.15
50	50	0.24	1.25	0.92	2.03	1.00	2.05	1.85	0.99	1.62
50	75	0.24	1.25	0.92	0.36	0.14	0.04	0.03	0.01	0.06
75	25	0.56	1.21	0.93	4.85	4.98	4.73	4.38	4.97	4.54
75	50	0.56	1.21	0.93	4.85	4.99	4.74	4.33	4.95	4.61
75	75	0.56	1.21	0.93	2.72	2.70	1.83	2.27	2.39	2.28

Note: The values displayed are the measured values multiplied by 10

Table 12. Marginal Cost of Major Therapeutic Procedures

Percentile		Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	2.64	2.57	1.97	0.05	0.03	0.04	0.03	0.10	0.09
25	50	2.96	2.91	2.18	2.79	1.78	2.02	2.36	1.73	1.12
25	75	3.82	3.86	2.92	4.99	4.97	4.97	4.99	4.89	4.84
50	25	2.64	2.57	1.97	0.09	0.04	0.10	0.01	0.03	0.09
50	50	2.96	2.91	2.18	0.95	1.73	0.34	1.73	1.59	0.76
50	75	3.82	3.86	2.92	4.99	4.94	4.97	4.99	4.89	4.84
75	25	2.64	2.57	1.97	0.03	0.08	0.03	0.01	0.01	0.07
75	50	2.96	2.91	2.18	0.03	0.11	0.07	0.03	0.03	0.20
75	75	3.82	3.86	2.92	4.40	4.03	4.60	4.24	3.71	4.80

Note: The values displayed are the measured values multiplied by 10

the noise. For all estimators, the marginal costs results are in line with the theory that marginal costs are increasing with scale. This property can be violated if using a non-parametric estimator without any shape constraints imposed. This can be seen in the marginal cost estimates for the Local Linear Kernel regression estimator in Appendix C, Table C.17 and Table C.18.

Our data set which combines AHA cost data with AHRQ output data for a broad sample of hospitals from across the US is unique to the best of our knowledge. However, the marginal cost estimates are broadly in line with marginal cost estimates for US hospitals for similar time periods. Gowrisankaran et al. (2015) studied considerable smaller set of hospitals located in Northern Virginia and observed in 2006 that on average are larger than the hospitals in our sample. Due to the difference in outputs the marginal cost levels are not directly comparable. However, conditional on the size variation, the variation in marginal costs is similar to the variation we observe for the

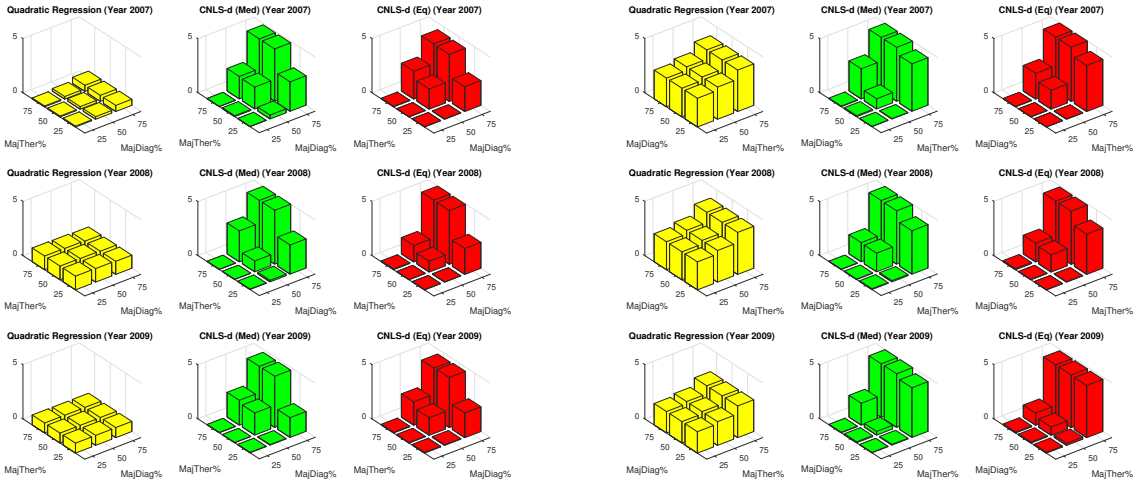

Figure 8. Marginal Cost of the Major Diagnostic Procedures (left) and Marginal Cost of the Major Therapeutic Procedures (right) ("CNLS-d (Med)" corresponds to CNLS-d using the median approach for the direction and "CNLS-d (Eq)" corresponds to CNLS-d using the direction with equal components in all netputs

parametric (quadratic) regression specification applied to our data. Boussemart et al. (2015) analyzed data on nearly 150 hospitals located in Florida observed in 2005. The authors use a different output specification and a trans-log model. The distribution of the size of the hospitals is similar to our dataset and we observe similar variances in marginal costs as the parametric (quadratic) regression specification applied to our data.

8. Conclusions

This paper investigated the improvement in functional estimates when specifying a particular direction in CNLS-d. From our Monte Carlo experiments, two primary findings emerged. First, directions close to the average orthogonal direction to the true function performed well. Second, when the data are noisy, selecting a direction that matched the noise direction of the DGP improves estimator performance. Our simulations indicate that CNLS-d with a direction orthogonal to the data is preferable if the noise level is not too large or a direction that matches the noise direction of the DGP if the noise level is large. Thus if the users know the shape of the data or the characteristics of the noise, they can use CNLS-d with a direction orthogonal to the data if the noise coefficient is small. Or if the noise coefficient is large, the user can select a direction close to the true noise direction, with non-zero components in all variables that potentially have noise. Our application to US hospital data shows that CNLS-d performs similarly across different directions which all include non-zero components of the direction vector for variables which potentially have noise in their measurement.

In future research, we propose developing an alternative estimator that incorporates multiple

directions in CNLS-d while maintaining the concavity axiom. This would allow dealing with subgroups within the data, permitting different assumptions across groups.

References

Ackerberg, D. A., Caves, K., Frazer, G., 2015. Identification properties of recent production function estimators. Econometrica 83 (6), 2411–2451.

Adler, N., Volta, N., 2016. Accounting for externalities and disposability: a directional economic environmental distance function. European Journal of Operational Research 250 (1), 314–327.

Afriat, S. N., 1972. Efficiency estimation of production functions. International Economic Review 13 (3), 568–598.

Aparicio, J., Pastor, J., Zofio, J., 2017. Can Farrell's allocative efficiency be generalized by the directional distance function approach? European Journal of Operational Research 257 (1), 345–351.

Atkinson, S., Cornwell, C., Honerkamp, O., 2003. Measuring and decomposing productivity change: stochastic distance function estimation versus data envelopment analysis. Journal of Business & Economic Statistics 21 (2), 284–294.

Atkinson, S., Tsionas, M., 2016. Directional distance functions: optimal endogenous directions. Journal of Econometrics 190 (2), 301–314.

Baležentis, T., De Witte, K., 2015. One- and multi-directional conditional efficiency measurement: efficiency in Lithuanian family farms. European Journal of Operational Research 245 (2), 612–622.

Bertsekas, D. P., 1999. Nonlinear programming. Athena Scientific, Belmont, MA.

Boussemart, J.-P., Leleu, H., Valdmanis, V., 2015. A two-stage translog marginal cost pricing approach for Floridian hospital outputs. Applied Economics 47 (38), 4116–4127.

Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., 2006. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Boca Raton, FL.

Chambers, R. G., 1988. Applied production analysis. Cambridge University Press, New York, NY.

Chambers, R. G., Chung, Y., Fare, R., 1996. Benefit and distance functions. Journal of Economic Theory 70 (2), 407–419.

Chambers, R. G., Chung, Y., Fare, R., 1998. Profit, directional distance functions, and nerlovian efficiency. Journal of Optimization Theory and Applications 98 (2), 351–364.

Charnes, A., Cooper, W. W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429–444.

Coelli, T., 2000. On the econometric estimation of the distance function representation of a production technology. Université Catholique de Louvain. Center for Operations Research and Econometrics [CORE].

Coelli, T., Perelman, S., 1999. A comparison of parametric and non-parametric distance functions: with application to European railways. European Journal of Operational Research 117 (2), 326–339.

Coelli, T., Perelman, S., 2000. Technical efficiency of European railways: a distance function approach. Applied Economics 32 (15), 1967–1976.

Diewert, W. E., Wales, T. J., 1987. Flexible functional forms and global curvature conditions. Econometrica 55 (1), 43–68.

Färe, R., Grosskopf, S., 2010. Directional distance functions and slacks-based measures of efficiency. European Journal of Operational Research 200 (1), 320–322.

Färe, R., Grosskopf, S., Noh, D.-W., Weber, W., 2005. Characteristics of a polluting technology: theory and practice. Journal of Econometrics 126 (2), 469–492.

Färe, R., Martins-Filho, C., Vardanyan, M., 2010. On functional form representation of multi-output production technologies. Journal of Productivity Analysis 33 (2), 81–96.

Färe, R., Pasurka, C., Vardanyan, M., 2016. On endogenizing direction vectors in parametric directional distance function-based models. European Journal of Operational Research 262 (1), 361–369.

Färe, R., Vardanyan, M., 2016. A note on parameterizing input distance functions: does the choice of a functional form matter? Journal of Productivity Analysis 45 (2), 121–130.

Ferrier, G. D., Leleu, H., Valdmanis, V. G., Vardanyan, M., 2018. A directional distance function approach for identifying the input/output status of medical residents. Applied Economics 50 (9), 1006–1021.
URL https://doi.org/10.1080/00036846.2017.1349287

Frisch, R., 1964. Theory of production. Springer Science & Business Media, Dordrecht, Netherlands.

Fukuyama, H., Matousek, R., 2018. Nerlovian revenue inefficiency in a bank production context: Evidence from shinkin banks. European Journal of Operational Research 271 (1), 317–330.

Gowrisankaran, G., Nevo, A., Town, R., 2015. Mergers when prices are negotiated: evidence from the hospital industry. American Economic Review 105 (1), 172–203.

Henderson, D. J., Parmeter, C. F., 2015. Applied Nonparametric Econometrics. Cambridge University Press, New York, NY.

Hildreth, C., 1954. Point estimates of ordinates of concave functions. Journal of the American Statistical Association 49 (267), 598–619.

Hollingsworth, B., 2003. Non-parametric and parametric applications measuring efficiency in health care. Health care management science 6 (4), 203–218.

Johnson, A. L., Kuosmanen, T., 2011. One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent stoNEZD method. Journal of Productivity Analysis 36 (2), 219–230.

Kapelko, M., Oude Lansink, A., 2017. Dynamic multi-directional inefficiency analysis of european dairy manufacturing firms. European Journal of Operational Research 257 (1), 338–344.

Koopmans, T. C., 1951. An analysis of production as an efficient combination of activities. Koopmans, T. C. (Ed.): Activity Analysis of Production and Allocation, Proceeding of a Conference, 33–97.

Kuosmanen, T., 2008. Representation theorem for convex nonparametric least squares. The Econometrics Journal 11 (2), 308–325.

Kuosmanen, T., Johnson, A., 2017. Modeling joint production of multiple outputs in stoned: Directional distance function approach. European Journal of Operational Research 262 (2), 792–801.

Kuosmanen, T., Johnson, A. L., 2010. Data envelopment analysis as nonparametric least-squares regression. Operations Research 58 (1), 149–160.

Kuosmanen, T., Kortelainen, M., 2012. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis 38 (1), 11–28.

Kutlu, L., 2018. A distribution-free stochastic frontier model with endogenous regressors. Economics Letters 163, 152–154.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. The Review of Economic Studies 70 (2), 317–341.

Lewbel, A., 2018. The identification zoo-meanings of identification in econometrics. Forthcoming in Journal of Economic Literature.

Li, Q., Racine, J. S., 2007. Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

Lim, E., 2014. On convergence rates of convex regression in multiple dimensions. INFORMS Journal on Computing 26 (3), 616–628.

Lovell, C. K., Travers, P., Richardson, S., Wood, L., 1994. Resources and functionings: a new view of inequality in Australia. In: Eichhorn (Ed.), Models and measurement of welfare and iInequality. Springer, Berlin, Germany, pp. 787–807.

Luenberger, D. G., 1992. Benefit functions and duality. Journal of Mathematical Economics 21 (5), 461–481.

Manski, C., 2003. Partial identification of probability distributions. Springer Series in Statistics. Springer, New York, NY.

Olley, G. S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–1297.

Pope, B., Johnson, A. L., 2013. Returns to scope: a metric for production synergies demonstrated for hospital production. Journal of Productivity Analysis 40 (2), 239–250.

Roshdi, I., Hasannasab, M., Margaritis, D., Rouse, P., 2018. Generalised weak disposability and efficiency measurement in environmental technologies. European Journal of Operational Research 266 (3), 1000–1012.

Shephard, R. W., 1953. Cost and production functions. Princeton University Press, Princeton, NJ.

Shephard, R. W., 1970. Theory of cost and production functions. Princeton University Press, Princeton, NJ.

Sickles, R. C., Good, D. H., Getachew, L., 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 17 (1-2), 133–155.

Syverson, C., 2011. What determines productivity? Journal of Economic Literature 49 (2), 326–365.

Tamer, E., 2010. Partial identification in econometrics. Annual Review of Economics 2, 167–195.

Varian, H. R., 1984. The nonparametric approach to production analysis. Econometrica, 579–597.

Wooldridge, J. M., 2009. On estimating firm-level production functions using proxy variables to control for unobservables. Economics Letters 104 (3), 112–114.

Yagi, D., Chen, Y., Johnson, A. L., Kuosmanen, T., 2018. Shape constrained kernel-weighted least squares: Estimating production functions for Chilean manufacturing industries. Journal of Business & Economic Statistics (forthcoming).

Zuckerman, S., Hadley, J., Iezzoni, L., 1994. Measuring hospital efficiency with frontier cost functions. Journal of Health Economics 13 (3), 255–280.

Appendices

The appendix is composed of the following parts:

- Properties of Directional Distance Functions and CNLS-d (Appendix A)
- Monte Carlo, Additional Experiments (Appendix B)
- Detailed results for the Hospital Application (Appendix C)

Appendix A. Properties of Directional Distance Functions and CNLS-d

Appendix A.1. Direction Selection in Directional Distance Functions

In this appendix we prove that the direction vector affects the functional estimates. Let $\boldsymbol{g}^{x,y} = (\boldsymbol{g}^x, \boldsymbol{g}^y)$ then we can state the following Theorem.

Theorem 1. *Suppose that two direction vectors exist, $\boldsymbol{g}_a^{x,y}$ and $\boldsymbol{g}_b^{x,y}$, such that $\boldsymbol{g}_a^{x,y} \neq \boldsymbol{g}_b^{x,y}$. Then the directional distance function estimates using these two different directions are not equal, $D(\boldsymbol{X}, \boldsymbol{Y}; \boldsymbol{g}_a^{x,y}) \neq D(\boldsymbol{X}, \boldsymbol{Y}; \boldsymbol{g}_b^{x,y})$.*

Proof. Rewrite Problem (10) as

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} (\alpha_i + \boldsymbol{\beta_i'}\, \boldsymbol{x_i} - {\gamma_i}'\, \boldsymbol{y_i})^2 \tag{A.1}$$

$$\text{s.t.} \quad \alpha_i + \boldsymbol{\beta_i'}\, \boldsymbol{x_i} - \boldsymbol{\gamma_i'}\, \boldsymbol{y_i} \leq \alpha_j + \boldsymbol{\beta_j'}\, \boldsymbol{x_i} - \boldsymbol{\gamma_j'}\, \boldsymbol{y_i}, \text{ for } i, j = 1, \ldots, n,\ i \neq j \tag{A.1a}$$

$$\boldsymbol{\beta_i}, \boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1, \ldots, n \tag{A.1b}$$

$$\boldsymbol{\beta_i'}\, \boldsymbol{g}^x + \boldsymbol{\gamma_i'}\, \boldsymbol{g}^y = 1, \text{ for } i = 1, \ldots, n \tag{A.1c}$$

Observe that all decision variables appear in the objective function and that the objective function is a quadratic function while the constraints define a convex solution space, i.e., this optimization problem has a unique solution (Bertsekas (1999)). If we solve Problem (A.1) with $\boldsymbol{g}_a^{x,y}$, then the resulting solution vector is $(\boldsymbol{\alpha}_a, \boldsymbol{\beta}_a, \boldsymbol{\gamma}_a)$. Changing the direction vector from $\boldsymbol{g}_a^{x,y}$ to $\boldsymbol{g}_b^{x,y}$ the normalization constraint $\boldsymbol{\beta_i'}\, \boldsymbol{g}_b^x + \boldsymbol{\gamma_i'}\, \boldsymbol{g}_b^y = 1$ no longer holds for $\boldsymbol{\beta}_a$ and $\boldsymbol{\gamma}_a$. However, the previous argument holds for the uniqueness of $(\boldsymbol{\alpha}_b, \boldsymbol{\beta}_b, \boldsymbol{\gamma}_b)$. Thus, $(\boldsymbol{\alpha}_a, \boldsymbol{\beta}_a, \boldsymbol{\gamma}_a) \neq (\boldsymbol{\alpha}_b, \boldsymbol{\beta}_b, \boldsymbol{\gamma}_b)$.

□

Appendix A.2. Details of CNLS-d

An alternative expression for CNLS-d (cf. Equations (16)-(16c)) is given by:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} \epsilon_i^2 \tag{A.2}$$

$$\text{s.t.} \quad -\epsilon_j + \epsilon_i + \boldsymbol{\beta_i'}\, (\boldsymbol{x_i} - \boldsymbol{x_j}) - \boldsymbol{\gamma_i'}\, (\boldsymbol{y_i} - \boldsymbol{y_j}) \leq 0, \text{ for } i, j = 1, \ldots, n,\ i \neq j \tag{A.2a}$$

$$\boldsymbol{\beta_i'}\, \boldsymbol{g}^x + \boldsymbol{\gamma_i'}\, \boldsymbol{g}^y = 1, \text{ for } i = 1, \ldots, n \tag{A.2b}$$

$$\boldsymbol{\beta_i},\ \boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1, \ldots, n. \tag{A.2c}$$

It is possible to recover $\alpha_i, i = 1, \ldots, n$, and the final estimates using the following relations:

$$\hat{\boldsymbol{x}}_{\boldsymbol{i}} = \boldsymbol{x}_{\boldsymbol{i}} + \epsilon_i \boldsymbol{g}^x, \text{ for } i = 1, \ldots, n \tag{A.3}$$

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{y}_{\boldsymbol{i}} - \epsilon_i \boldsymbol{g}^y, \text{ for } i = 1, \ldots, n \tag{A.4}$$

$$\alpha_i = -\boldsymbol{\beta}_{\boldsymbol{i}}' \boldsymbol{x}_{\boldsymbol{i}} + \boldsymbol{\gamma}_{\boldsymbol{i}}' \boldsymbol{y}_{\boldsymbol{i}} + \epsilon_i, \text{ for } i = 1, \ldots, n. \tag{A.5}$$

Appendix A.3. Different Directions for Different Groups in CNLS-d

Consider the case where all observations have the same input level and produce two outputs and estimate the isoquant. Define two groups of observations G_1 and G_2 such that $|G_1 \cup G_2| = n$ and $G_1 \cap G_2 = \emptyset$.[15] Using the notation in Section Appendix A.1, the direction vector for the first group of observations G_1 is $\boldsymbol{g}^{y_{G_1}}$ and it's $\boldsymbol{g}^{y_{G_2}}$ for the second group of observations G_2.

For either a fixed input vector, $\boldsymbol{X}$, or a fixed cost level, c, formulate the iso-cost estimator for G_1 and G_2 with different directions vectors as:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{A.6}$$

$$\text{s.t.} \quad -\epsilon_j + \epsilon_i - \boldsymbol{\gamma}_{\boldsymbol{i}}' (\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{y}_{\boldsymbol{j}}) \leq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j \tag{A.6a}$$

$$\boldsymbol{\gamma}_{\boldsymbol{i}}' \boldsymbol{g}^{\boldsymbol{y_{G_1}}} = 1, \text{ for } i \in G_1 \tag{A.6b}$$

$$\boldsymbol{\gamma}_{\boldsymbol{i}}' \boldsymbol{g}^{\boldsymbol{y_{G_2}}} = 1, \text{ for } i \in G_2 \tag{A.6c}$$

$$\boldsymbol{\gamma}_{\boldsymbol{i}} \geq 0, \text{ for } i = 1, \ldots, n. \tag{A.6d}$$

Note that using more than one direction for CNLS-d can lead to violations on convexity. Only under very limiting conditions can we allow for multiple directions in CNLS-d and guarantee that the resulting estimated function will maintain convexity. The following theorem formalizes the conditions.

Theorem 2. *If a CNLS-d estimator is calculated using two groups of observations with different direction vectors as shown in Equation* (A.6)*, and the following condition holds regarding the direction vectors and the noise direction:*

$$\left(\epsilon_i \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} \right)' \left[\frac{\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}\|} - \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} \right] \geq 0, \ \textit{for } i, j = 1, \ldots, n, \ i \neq j, \tag{A.7}$$

[15] The notation $|\cdot|$ corresponds to the cardinality of the set.

where

$$k(i) = \begin{cases} 1, & \textit{if } i \in G_1 \\ 2, & \textit{if } i \in G_2, \end{cases}$$

then, the resulting CNLS-d estimate is a concave function.

Proof. Consider the Afriat inequalities in the context of cost isoquant estimation. One of the conditions of Equation (16) is:

$$\epsilon_i - \epsilon_j - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \leq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j. \tag{A.8}$$

Knowing that $\epsilon_i \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} = \hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}$ means that $\epsilon_i = \left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right)' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|}$.

Substituting ϵ_i and ϵ_j in the inequalities (A.8) obtains:

$$\left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right)' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} - \left(\hat{\boldsymbol{y}}_{\boldsymbol{j}} - \boldsymbol{y_j}\right)' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}\|} - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \leq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j. \tag{A.9}$$

Next, consider the case where both observations have the same direction. Then the expression is:

$$\left[\left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right) - \left(\hat{\boldsymbol{y}}_{\boldsymbol{j}} - \boldsymbol{y_j}\right)\right]' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \leq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j. \tag{A.10}$$

If Equation (A.10) is satisfied, we know that the CNLS-d constraints hold. By comparison observe that the condition listed below is a sufficient condition for Equation (A.10) being satisfied when Equation (A.9) holds:

$$\begin{aligned} &\left[\left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right) - \left(\hat{\boldsymbol{y}}_{\boldsymbol{j}} - \boldsymbol{y_j}\right)\right]' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \quad -- \textit{from eq.}(\text{A.10}) \\ \leq &\left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right)' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|} - \left(\hat{\boldsymbol{y}}_{\boldsymbol{j}} - \boldsymbol{y_j}\right)' \frac{\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}\|} - \boldsymbol{\gamma_i'} \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \quad -- \textit{from eq.}(\text{A.9}) \\ &\text{for } i, j = 1, \ldots, n, \ i \neq j, \end{aligned}$$

which, after simplifying, becomes:

$$\left(\hat{\boldsymbol{y}}_{\boldsymbol{i}} - \boldsymbol{y_i}\right)' \left[\frac{\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(j)}}}\|} - \frac{\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}}{\|\boldsymbol{g}^{\boldsymbol{y_{k(i)}}}\|}\right] \geq 0 \text{ for } i, j = 1, \ldots, n, \ i \neq j \tag{A.11}$$

□

Thus Theorem 2 is proved and a sufficient condition is found that, if verified, ensures the concavity property of the estimator even when multiple directions are used in the estimation of the directional distance function.

The following corollary, concerning the convex case, is directly infered from Theorem 2:

Corollary 1. *If a CNLS-d estimator is calculated using two groups of observations with different direction vectors as shown in Equation* (A.6)*, and the following condition holds regarding the direction vectors and the noise direction:*

$$\left(\epsilon_i \frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}\|}\right)' \left[\frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(j)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(j)}}\|} - \frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}\|}\right] \leq 0, \text{ for } i = 1, \ldots, n, \ i \neq j, \tag{A.12}$$

where

$$k(i) = \begin{cases} 1, & \text{if } i \in G_1 \\ 2, & \text{if } i \in G_2, \end{cases}$$

then, the resulting CNLS-d estimate is a convex function.

Proof. Invert the sign from Equation (A.8):

$$-\epsilon_j + \epsilon_i - \boldsymbol{\gamma}_{\boldsymbol{i}}' \left(\boldsymbol{y_i} - \boldsymbol{y_j}\right) \geq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j, \tag{A.13}$$

and follow the logic of the proof of Theorem 2 to obtain Corollary 1 and Equation (A.12).

□

Theorem 2 clarifies that if the directions for each respective group are orthogonal to each other, then condition A.7 is verified. This means that if the direction for group 1 has a single nonzero component in the output 1 dimension and group 2 has a single nonzero component in the output 2 dimension, then we will not observe violations of the convexity property.

We state a second Corollary that follows from Theorem 2 and is useful when there are more than two groups each with their own estimation direction in CNLS-d.

Corollary 2. *Let $n \in \mathbb{N}$ the total number of observation. Let Q the number of outputs considered. Let $\boldsymbol{Y} = \{\boldsymbol{y}_i \in \mathbb{R}_+^Q, i = 1, \ldots, n\}$ the set of observed outputs. Let P_g a partition of $\boldsymbol{Y}$ of cardinal $N_g \in \mathbb{N}$. Let $\boldsymbol{g}^y = \{\boldsymbol{g}^{y_k}, k = 1, \ldots, N_g\}$ the set of directions used for each respective group of the partition. If a CNLS-d estimator is calculated using the directions from $\boldsymbol{g}^y$ based on partition P_g, and the following condition holds regarding the direction vectors and the noise direction:*

$$\left(\epsilon_i \frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}\|}\right)' \left[\frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(j)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(j)}}\|} - \frac{\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}}{\|\boldsymbol{g}^{\boldsymbol{y}_{\boldsymbol{k}(i)}}\|}\right] \geq 0, \text{ for } i, j = 1, \ldots, n, \ i \neq j, \tag{A.14}$$

where for each $i = 1, \ldots, n$, $k(i)$ corresponds to the indicator of the part of the partition P_g, in which $\boldsymbol{y}_i$ belongs. Then the resulting CNLS-d estimate is a concave function.

Proof. We can follow the proof of Theorem 2, as the condition does not change. The condition still concerns observations pairwise, the only difference is that now the partition of observations corresponds to more than two groups. This does not affect the proof of the condition.

□

Corollary 2 extends the statement of Theorem 2 to provide sufficient conditions to avoid violations of the shape constraints in a scenario where there are more than two groups each with their own estimation direction in CNLS-d estimation.

Simulations to investigate the frequency in which multiple directions leads to violations. We run simulations to investigate the effects of using multiple directions. We use the same DGP as stated in Section 5, Example 1. However, we define two groups and assign different directions for each one of them:

$$G_1 = \{i \in \{1,2,...,n\} | \arctan(\tilde{y}_{i2}/\tilde{y}_{i1}) \leq \pi/4\} \tag{A.15}$$

$$G_2 = \{i \in \{1,2,...,n\} | \arctan(\tilde{y}_{i2}/\tilde{y}_{i1}) > \pi/4\}, \tag{A.16}$$

and,

$$\boldsymbol{g^y} = \begin{cases} \boldsymbol{g^{y_{G_1}}}, & \text{if } i \in G_1 \\ \boldsymbol{g^{y_{G_2}}}, & \text{if } i \in G_2, \end{cases} \tag{A.17}$$

where $\boldsymbol{g^{y_{G_1}}} = [\cos(\pi/8), \sin(\pi/8)]$ and $\boldsymbol{g^{y_{G_2}}} = [\cos(3\pi/8), \sin(3\pi/8)]$.

We run a total of 100 simulations. For comparison, for each simulation, we also record the estimates when using only the direction based on $\pi/8$ and $3\pi/8$ only for all observations. We identify violations of the monotonicity and concavity by sorting the estimates by y_1. We identify all adjacent pairs and triplets, which means 99 pairs and 98 triplets given that we consider 100 observations for each simulation.

As expected, there are no violations when we use a single direction for the estimation. However, when we use two directions we do observe violations. For monotonicity, we observe no violations for pairs of observations that are part of the same group. However, for pairs with one member from each group we observe violations of monotonicity for 6% of the pairs. We use the triplets to analyze concavity. When the members of the triplet are from the same group, we observe violations of concavity for 2% of the triplets. When one member of the triplet is from a different group, the violations of concavity increase to 45%. These results indicate that for one instance when the conditions of Theorem 2 do not hold, we see a significant number of violations of the imposed assumptions.

Appendix B. Additional Experiments

Measuring MSE Example, Section 4.1 - Noise Generated in a Common and Prespecified Direction θ_f.

This section describes the simulations and the results for the fixed noise direction case referenced in Section 4.1.

The Data Generation Process (DGP) for observations $(\boldsymbol{y}_i, c_i), i = 1, \ldots, n$, is as follows:

1. The output, $\tilde{y}_i$, is drawn from the continuous uniform distribution $U[0, 1]$
2. The cost is calculated as $\tilde{c}_i = \beta_0\ \tilde{y}_i$, where $\beta_0 = 1$.
3. In the case of fixed direction, the noise term is determined as:
   (a) l_{ϵ_i} is the scalar length that is drawn from a normal distribution, $N(0, \lambda\,\epsilon_0)$, λ is prespecified and an initial value for the standard deviation, ϵ_0, is calculated as in Equation (11) in Section 4.1.:

$$\epsilon_0 = \frac{1}{2}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\tilde{y}_i - \bar{y})^2} + \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\tilde{c}_i - \bar{c})^2}\right], \tag{B.1}$$

   where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i$ and $\bar{c} = \frac{1}{n}\sum_{i=1}^{n}\tilde{c}_i$ are the mean of the output and the mean of the cost without noise, respectively.
   (b) $\boldsymbol{v_f} = [\cos(\theta_f), \sin(\theta_f)]$ is the fixed noise direction that is inferred from the prespecified angle θ_f.
   (c) $(\epsilon_{y_i}, \epsilon_{c_i}) = l_{\epsilon_i}\,\boldsymbol{v_f},\ i = 1, \ldots, n.$
4. The observations with noise are obtained by adding the noise term:

$$\begin{pmatrix} y_i \\ c_i \end{pmatrix} = \begin{pmatrix} \tilde{y}_i \\ \tilde{c}_i \end{pmatrix} + \begin{pmatrix} \epsilon_{y_i} \\ \epsilon_{c_i} \end{pmatrix},\ i = 1, \ldots, n. \tag{B.2}$$

Figure B.9. Linear function data generation process with fixed noise direction

Apply the DGP described above to generate a training set, $(y_{tr_i}, c_{tr_i}),\ i = 1, \ldots, n_{tr}$, and a testing set $(y_{ts_i}, c_{ts_i}),\ i = 1, \ldots, n_{ts}$. Consider 100 repetitions of the simulation and set the number of observations in each group to $n_{tr} = n_{ts} = 100$. Set the scaling coefficient for the noise to $\lambda = 0.6$. Consider different DGP since data is generated for the following values of noise direction angles, $\theta_f \in \{0, \pi/8, \pi/4, 3\pi/8, \pi/2\}$.

We test the set of directions corresponding to the angle $\theta_t \in \{0, \pi/8, \pi/4, 3\pi/8, \pi/2\}$. If the

direction of the noise, θ_f, matches the direction used in the DDF, θ_t, then the smallest MSE results for all cases.

Results: Fixed Noise Direction. Table B.13 reports the MSE computed by comparing the estimated function to the true function and Table B.14 reports the MSE computed by comparing the estimated function to the testing set.

Table B.13. Average MSE over 100 simulations for the Linear Estimator compared to the true function with a DGP using random noise directions

		Average MSE: Estimator compared to the true function DDF Direction Angle θ_t				
Noise Dir Angle θ_f	MSE Dir Ang θ_{MSE}	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	0	**0.55**	1.59	3.49	6.35	12.06
0	$\pi/8$	**0.32**	0.86	1.81	3.17	5.70
0	$\pi/4$	**0.27**	0.69	1.42	2.44	4.23
0	$3\pi/8$	**0.32**	0.77	1.54	2.58	4.36
0	$\pi/2$	**0.54**	1.21	2.37	3.86	6.28
$\pi/8$	0	3.22	**1.00**	2.66	7.79	22.92
$\pi/8$	$\pi/8$	2.16	**0.59**	1.39	3.80	9.98
$\pi/8$	$\pi/4$	2.04	**0.50**	1.10	2.88	7.09
$\pi/8$	$3\pi/8$	2.67	**0.59**	1.21	3.02	7.02
$\pi/8$	$\pi/2$	5.40	**1.03**	1.88	4.45	9.68
$\pi/4$	0	8.95	2.92	**1.18**	2.95	15.94
$\pi/4$	$\pi/8$	6.46	1.93	**0.70**	1.53	7.21
$\pi/4$	$\pi/4$	6.49	1.81	**0.61**	1.20	5.24
$\pi/4$	$3\pi/8$	9.10	2.35	**0.74**	1.31	5.30
$\pi/4$	$\pi/2$	20.84	4.70	**1.32**	2.03	7.48
$3\pi/8$	0	9.65	4.44	1.90	**1.13**	5.70
$3\pi/8$	$\pi/8$	6.99	3.00	1.22	**0.65**	2.83
$3\pi/8$	$\pi/4$	7.05	2.86	1.11	**0.55**	2.17
$3\pi/8$	$3\pi/8$	9.92	3.76	1.40	**0.64**	2.30
$3\pi/8$	$\pi/2$	22.76	7.71	2.66	**1.09**	3.45
$\pi/2$	0	6.15	3.76	2.29	1.16	**0.50**
$\pi/2$	$\pi/8$	4.25	2.50	1.49	0.73	**0.29**
$\pi/2$	$\pi/4$	4.11	2.36	1.37	0.66	**0.25**
$\pi/2$	$3\pi/8$	5.52	3.06	1.74	0.81	**0.29**
$\pi/2$	$\pi/27$	11.62	6.10	3.33	1.50	**0.49**

Note: Displayed are the measured values multiplied by 10^3

In Table B.13, the direction for the DDF corresponding to the smallest MSE always matches the noise direction in the DGP. Further for more than 70% of the cases tested there is more than a

Table B.14. Average MSE over 100 simulations for the Linear Estimator compared to an out-of-sample testing set with a DGP using fixed noise directions

		Average MSE: Estimator compared to testing set data DDF Direction Angle θ_t				
Noise Dir Angle θ_f	MSE Dir Ang θ_{MSE}	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	0	**30.02**	31.22	33.23	36.21	42.08
0	$\pi/8$	17.53	**17.13**	17.46	18.24	20.01
0	$\pi/4$	14.95	13.99	**13.86**	14.10	14.92
0	$3\pi/8$	17.51	15.70	15.15	**15.03**	15.42
0	$\pi/2$	29.93	25.30	23.55	22.64	**22.32**
$\pi/8$	0	**49.89**	52.78	58.59	68.39	91.28
$\pi/8$	$\pi/8$	32.41	**30.88**	31.71	34.14	40.37
$\pi/8$	$\pi/4$	29.93	26.38	**25.69**	26.27	28.92
$\pi/8$	$3\pi/8$	38.15	31.00	28.66	**27.92**	28.88
$\pi/8$	$\pi/2$	74.19	53.30	45.83	41.93	**40.19**
$\pi/4$	0	**51.54**	53.79	59.55	70.76	101.99
$\pi/4$	$\pi/8$	36.65	**34.53**	35.21	38.14	47.22
$\pi/4$	$\pi/4$	36.39	31.60	**30.32**	30.83	34.75
$\pi/4$	$3\pi/8$	50.32	39.87	35.91	**34.32**	35.52
$\pi/4$	$\pi/2$	112.21	76.31	62.47	54.76	**50.83**
$3\pi/8$	0	**39.37**	41.09	45.01	52.54	73.64
$3\pi/8$	$\pi/8$	28.28	**27.35**	28.14	30.56	37.89
$3\pi/8$	$\pi/4$	28.30	25.72	**25.22**	26.01	29.73
$3\pi/8$	$3\pi/8$	39.47	33.40	31.11	**30.42**	32.19
$3\pi/8$	$\pi/2$	89.14	66.84	57.41	51.96	**49.51**
$\pi/2$	0	**22.47**	22.94	23.97	25.85	30.66
$\pi/2$	$\pi/8$	15.44	**15.16**	15.36	15.99	17.91
$\pi/2$	$\pi/4$	14.89	14.17	**14.01**	14.21	15.27
$\pi/2$	$3\pi/8$	19.88	18.27	17.59	**17.35**	17.88
$\pi/2$	$\pi/2$	41.52	36.04	33.31	31.51	**30.54**

Note: Displayed are the measured values multiplied by 10^3

50% decrease in MSE by using the correctly specified direction compared to the next best direction tested, which was not as large in the random direction case in Table 1 of Section 4.1. In other words, when endogeneity is severe, the benefits of using a DDF with a well-selected direction are potentially large.

Table B.14 is consistent with the results observed in the random noise case, in Table 2 of Section 4.1. The DDF directions corresponding to the smallest MSE values are those matching the directions used for the MSE computation. Thus, the proposed radial MSE measure addresses the

challenge of measuring performance in applications with a testing dataset.

Monte Carlo Simulations - Experiments, Section 5.2 - Experiment 3. Base case with fixed noise direction and different noise levels.

This section summarizes results Experiment 3 with $lambda = 0.2$.

Table B.15. Experiment 3–More Noise: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.2.

	CNLS-d Direction Angle				
Noise Direction Angle	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0	**8.15**	15.62	37.66	82.16	183.39
$\pi/8$	50.60	**11.59**	20.68	67.88	206.46
$\pi/4$	145.21	29.40	**11.89**	33.89	149.24
$3\pi/8$	220.24	69.87	22.28	**11.66**	53.66
$\pi/2$	165.84	72.13	33.27	14.25	**7.41**

Note: Displayed are measured values multiplied by 10^4

Appendix C. U.S. Hospital Dataset Application

We describe the functional estimates provided by Quadratic Regression, CNLS-d using the median direction, and Local Linear Kernel. Table C.16 provides most productive scale size (MPSS) measurements in units of normalized cost. The values reported in Table C.16 are normalized costs on the scale of 0 to 1. Tables C.17 and C.18 provide marginal cost of Major Diagnostic procedures and marginal cost of Major Therapeutic procedures, respectively. The units for Tables C.17 and C.18 are normalized cost over normalized Major Diagnostic procedures and Major Therapeutic procedures, respectively.

Our conclusions are the same as stated in the body of the paper, CNLS-d provides the advantage of being more flexible than the parametric estimator (Quadratic regression), while having shape constraints that maintain the interpretability of the results. The interpretability issue for a nonparametric estimator without any axiomatic shape constraints imposed is illustrated by the LL Kernel estimates, for which we observe several negative marginal costs for Major Diagnostic procedures in 2008 and 2009. These negative values often appear in consecutive 25 and 50 percentiles. This is because most hospitals are small, which implies the corresponding output levels are close for 25 and 50 percentiles and the LL Kernel estimator uses relatively large bandwidth values making estimates relatively smooth. Thus, if there is a negative marginal cost for 25 percentile estimate, then the 50th percentile will often be negative too.

Table C.16. Most Productive Scale Size (c)

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	2.84	0.84	3.66	0.40	0.30	0.35	8.40	2.24	6.69	1.41	1.11	3.60
25	25	25	50	0.46	0.42	0.39	2.01	0.45	1.45	4.91	1.57	1.76	0.99	0.69	0.93
25	25	25	75	0.42	0.39	0.32	7.04	0.46	3.48	8.38	1.47	1.56	0.93	0.60	0.78
25	25	50	25	0.40	2.93	2.30	0.97	0.45	0.98	2.53	3.60	3.05	1.22	0.53	0.79
25	25	50	50	0.69	0.80	0.73	0.43	0.39	0.42	2.29	2.11	2.25	1.18	0.93	1.10
25	25	50	75	0.40	0.41	0.33	1.00	0.44	1.47	3.22	1.52	1.67	1.42	0.66	0.89
25	25	75	25	0.26	1.73	1.34	0.99	0.47	2.53	1.99	2.42	2.67	0.91	0.57	0.11
25	25	75	50	0.41	1.01	0.70	0.95	0.45	1.01	2.95	2.18	1.97	1.66	0.54	0.75
25	25	75	75	0.66	0.38	0.63	0.44	0.39	0.43	2.28	2.03	2.05	1.18	0.87	0.99
25	50	25	25	0.36	0.16	0.30	2.02	0.40	0.45	4.33	3.58	5.51	2.34	1.14	4.46
25	50	25	50	6.18	0.81	5.60	0.46	0.33	0.37	12.62	2.42	6.92	5.90	1.31	0.82
25	50	25	75	1.00	0.43	0.41	1.51	0.43	1.43	3.98	1.55	1.73	1.00	0.69	0.88
25	50	50	25	1.40	1.39	1.62	0.45	0.36	0.42	4.31	3.96	3.99	2.14	1.25	1.46
25	50	50	50	1.79	1.21	5.67	0.39	0.30	0.35	7.98	2.05	10.91	5.03	0.98	4.74
25	50	50	75	0.96	0.45	0.42	0.97	0.42	0.94	3.43	1.95	1.85	1.52	1.08	1.00
25	50	75	25	0.69	2.07	3.01	0.97	0.44	0.50	2.70	3.20	3.98	1.13	0.73	0.33
25	50	75	50	0.59	2.90	4.48	0.92	0.42	0.47	3.25	3.03	6.17	1.82	0.71	1.09
25	50	75	75	0.79	0.84	0.79	0.43	0.38	0.42	2.49	2.15	2.33	1.29	0.96	1.14
25	75	25	25	0.40	0.22	0.39	5.03	0.41	0.47	7.39	2.07	4.48	1.37	0.40	3.86
25	75	25	50	0.11	0.97	0.65	1.01	0.39	0.45	4.33	4.97	6.13	3.05	1.83	4.74
25	75	25	75	6.55	0.80	6.48	0.46	0.33	0.38	12.90	2.40	5.22	5.83	1.30	0.76
25	75	50	25	0.30	0.08	0.23	1.04	0.40	0.46	2.56	2.42	4.20	1.60	0.44	3.72
25	75	50	50	0.12	1.31	0.94	0.98	0.39	0.44	4.26	5.35	6.25	2.95	1.96	4.58
25	75	50	75	6.64	0.83	6.56	0.45	0.32	0.37	13.24	2.43	4.95	6.10	1.31	0.81
25	75	75	25	0.34	0.92	1.18	0.46	0.36	0.43	2.83	2.99	3.06	1.67	0.83	1.18
25	75	75	50	1.73	2.04	2.05	0.45	0.35	0.42	5.19	5.04	4.33	2.68	1.72	1.84
25	75	75	75	1.84	1.22	6.43	0.39	0.30	0.35	8.13	2.05	4.49	5.55	0.97	0.91
50	25	25	25	0.86	0.89	0.57	1.01	0.41	0.77	2.70	2.56	1.88	0.79	0.59	0.56
50	25	25	50	0.46	0.41	0.33	0.52	0.43	0.45	3.33	1.22	1.66	1.39	0.35	0.88
50	25	25	75	0.42	0.39	0.30	1.52	0.45	1.98	3.37	1.50	1.56	1.41	0.63	0.79
50	25	50	25	0.43	3.13	0.72	0.49	0.46	1.02	1.50	4.05	2.15	0.79	0.83	0.38
50	25	50	50	0.70	0.79	0.64	0.42	0.37	0.40	2.66	2.22	1.76	1.53	1.05	1.03
50	25	50	75	0.41	0.41	0.31	1.00	0.44	0.98	3.24	1.55	1.68	1.43	0.69	0.91
50	25	75	25	0.28	1.81	0.93	0.99	0.47	2.05	1.64	2.54	2.31	0.88	0.57	0.11
50	25	75	50	0.42	1.02	0.66	0.95	0.45	1.00	1.59	2.05	1.80	0.73	0.38	0.60
50	25	75	75	0.66	0.39	0.61	0.44	0.39	0.43	2.29	2.08	2.04	1.19	0.92	0.99
50	50	25	25	0.05	0.59	0.89	0.52	0.38	0.43	3.33	4.24	5.87	2.72	1.52	4.34
50	50	25	50	6.65	0.80	5.54	0.45	0.32	0.36	13.77	2.55	7.13	6.52	1.44	1.14
50	50	25	75	1.01	0.43	0.40	1.01	0.43	0.95	4.01	1.59	1.72	1.51	0.72	0.88
50	50	50	25	2.03	1.63	1.90	0.45	0.34	0.40	5.27	4.21	3.29	2.48	1.32	0.80
50	50	50	50	1.82	1.21	5.58	0.38	0.29	0.34	3.54	2.20	8.63	1.35	1.12	2.58
50	50	50	75	0.97	0.45	0.41	0.97	0.42	0.94	3.46	2.02	1.84	1.54	1.15	1.00
50	50	75	25	0.70	2.14	2.94	0.48	0.44	0.50	2.03	3.16	3.74	1.25	0.71	0.14
50	50	75	50	0.60	2.63	4.38	0.92	0.42	0.47	3.38	3.83	5.23	1.90	1.48	0.88
50	50	75	75	0.80	0.84	0.77	0.43	0.37	0.42	2.51	2.20	2.30	1.30	1.01	1.13
50	75	25	25	0.04	0.05	0.36	1.54	0.41	0.46	3.31	2.28	4.52	1.66	0.47	3.91
50	75	25	50	0.23	1.03	0.62	1.01	0.39	0.45	4.28	5.19	6.14	3.33	1.93	4.76
50	75	25	75	6.64	0.80	6.39	0.46	0.33	0.38	13.21	2.45	4.69	6.03	1.34	1.11
50	75	50	25	0.38	0.14	0.21	1.03	0.40	0.45	2.93	2.67	4.23	1.86	0.56	3.75
50	75	50	50	0.24	1.37	0.90	0.98	0.38	0.44	4.22	5.62	6.22	3.20	2.12	2.93
50	75	50	75	6.73	0.83	6.48	0.44	0.32	0.37	13.60	2.48	4.85	6.35	1.35	1.23
50	75	75	25	0.47	0.97	1.13	0.46	0.36	0.43	3.16	3.14	2.86	1.87	0.82	1.02
50	75	75	50	1.84	2.09	2.00	0.45	0.35	0.42	5.52	5.01	4.03	2.88	1.82	1.33
50	75	75	75	1.85	1.23	6.36	0.39	0.30	0.35	8.01	2.10	8.51	5.35	1.01	4.95
75	25	25	25	0.55	0.42	0.08	3.14	1.43	0.88	4.32	2.72	2.16	0.06	0.11	0.82
75	25	25	50	0.38	0.52	0.15	0.54	0.46	0.86	2.03	1.23	1.47	0.64	0.23	0.36
75	25	25	75	0.42	0.37	0.25	0.53	0.45	0.46	2.18	1.14	1.43	1.20	0.30	0.72
75	25	50	25	0.47	0.52	0.20	1.14	0.47	0.93	2.35	2.64	2.56	0.29	0.28	0.69
75	25	50	50	0.67	0.56	0.34	0.50	0.43	0.85	2.38	1.56	2.41	0.67	0.52	0.41
75	25	50	75	0.41	0.38	0.26	0.51	0.44	0.45	3.05	1.17	1.57	1.68	0.33	0.85
75	25	75	25	0.32	2.25	0.23	0.53	0.50	1.08	1.71	2.69	2.11	0.35	0.45	0.09
75	25	75	50	0.45	0.97	0.33	0.50	0.47	1.04	2.06	2.30	1.92	0.61	0.59	0.66
75	25	75	75	0.68	0.74	0.53	0.43	0.36	0.41	2.60	1.46	1.59	1.47	0.33	0.62
75	50	25	25	1.18	2.50	0.49	1.08	0.85	0.81	2.88	5.64	4.18	0.59	2.28	2.83
75	50	25	50	0.97	0.93	0.59	0.99	0.40	0.76	2.93	2.32	1.84	0.93	0.62	0.55
75	50	25	75	0.99	0.41	0.34	0.52	0.43	0.44	2.86	1.21	1.94	1.31	0.36	0.81
75	50	50	25	1.07	2.83	0.51	0.52	0.41	0.42	2.82	3.96	3.36	0.70	1.24	1.73
75	50	50	50	0.88	0.98	0.61	0.93	0.38	0.77	2.82	2.02	1.99	0.97	0.66	0.60
75	50	50	75	0.96	0.42	0.35	0.50	0.41	0.43	3.32	1.68	2.09	1.84	0.39	0.95
75	50	75	25	0.50	2.56	0.69	0.51	0.47	1.02	2.31	3.41	2.75	0.79	0.74	0.27
75	50	75	50	0.62	1.42	0.81	0.49	0.45	0.98	2.41	4.03	2.40	1.08	1.15	0.72
75	50	75	75	0.81	0.81	0.68	0.43	0.35	0.40	2.86	2.30	1.81	1.62	1.13	1.06
75	75	25	25	0.19	0.22	0.62	0.53	0.37	0.43	2.63	3.04	4.82	2.23	0.91	3.67
75	75	25	50	0.26	1.57	1.60	0.51	0.36	0.42	4.41	5.73	6.57	3.56	2.33	4.43
75	75	25	75	4.16	0.78	1.73	0.44	0.32	0.36	4.95	2.57	3.10	1.37	1.47	1.02
75	75	50	25	0.19	0.56	0.87	0.51	0.37	0.42	2.85	3.56	4.98	2.25	1.18	3.64
75	75	50	50	0.42	1.86	1.81	0.48	0.36	0.41	4.40	6.27	6.70	3.68	2.66	4.41
75	75	50	75	3.09	0.81	1.75	0.43	0.31	0.36	4.44	2.65	3.22	1.42	1.54	1.12
75	75	75	25	1.68	1.38	1.75	0.45	0.34	0.40	4.53	3.39	3.14	2.10	0.98	0.89
75	75	75	50	2.90	2.39	2.50	0.43	0.33	0.39	6.97	5.12	3.81	3.33	1.84	0.75
75	75	75	75	1.42	1.21	2.16	0.38	0.29	0.34	3.16	2.26	5.67	1.36	1.18	3.11

Note: The values displayed are the measured values multiplied by 10

Table C.17. Marginal Cost of Major Diagnostic Procedures

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	0.08	1.26	0.92	0.63	0.49	0.24	1.52	3.14	2.58	1.06	1.48	0.91
25	25	25	50	0.08	1.26	0.92	0.02	0.03	0.03	0.98	2.78	1.62	0.81	1.67	0.74
25	25	25	75	0.08	1.26	0.92	0.25	0.11	0.04	0.93	3.19	2.73	0.79	2.12	1.69
25	25	50	25	0.24	1.25	0.92	3.80	4.27	4.13	4.74	5.34	5.36	0.94	0.94	0.64
25	25	50	50	0.24	1.25	0.92	2.41	2.36	2.45	2.71	3.35	3.59	0.79	1.11	0.39
25	25	50	75	0.24	1.25	0.92	0.41	0.26	0.04	1.09	2.74	2.50	0.83	1.62	1.41
25	25	75	25	0.56	1.21	0.93	4.77	4.87	4.99	4.99	6.13	6.67	0.48	0.15	0.47
25	25	75	50	0.56	1.21	0.93	4.84	4.98	4.91	5.01	6.14	6.28	0.49	0.16	0.34
25	25	75	75	0.56	1.21	0.93	2.93	2.70	1.90	3.49	3.93	3.47	0.63	0.75	0.02
25	50	25	25	0.08	1.26	0.92	0.14	0.01	0.08	1.30	2.72	2.85	1.14	1.62	1.64
25	50	25	50	0.08	1.26	0.92	0.02	0.02	0.02	1.09	2.93	2.52	0.96	1.88	1.56
25	50	25	75	0.08	1.26	0.92	0.02	0.06	0.04	1.08	3.23	3.03	0.97	2.15	1.98
25	50	50	25	0.24	1.25	0.92	2.09	2.27	2.68	3.83	3.57	4.78	1.09	0.92	1.34
25	50	50	50	0.24	1.25	0.92	1.84	1.29	2.71	2.37	3.05	3.49	0.98	1.14	1.13
25	50	50	75	0.24	1.25	0.92	0.36	0.26	0.04	1.24	2.74	2.65	0.99	1.73	1.59
25	50	75	25	0.56	1.21	0.93	4.85	4.98	4.77	4.89	5.72	6.75	0.50	-0.25	1.02
25	50	75	50	0.56	1.21	0.93	4.85	4.99	4.83	4.93	5.77	6.70	0.52	-0.19	0.75
25	50	75	75	0.56	1.21	0.93	2.93	2.70	1.83	3.46	3.65	3.89	0.76	0.44	0.57
25	75	25	25	0.08	1.26	0.92	0.07	0.02	0.01	1.53	2.97	4.47	1.44	1.85	3.42
25	75	25	50	0.08	1.26	0.92	0.01	0.02	0.02	1.31	2.86	4.46	1.18	1.79	3.59
25	75	25	75	0.08	1.26	0.92	0.02	0.02	0.04	1.23	2.79	3.54	1.02	1.57	2.61
25	75	50	25	0.24	1.25	0.92	0.38	0.08	0.04	2.21	2.20	4.65	1.25	1.01	3.01
25	75	50	50	0.24	1.25	0.92	0.38	0.11	0.08	1.46	2.22	4.83	1.20	1.12	3.33
25	75	50	75	0.24	1.25	0.92	0.02	0.16	0.04	1.40	2.13	3.56	1.15	1.32	2.78
25	75	75	25	0.56	1.21	0.93	3.16	3.05	2.38	4.42	3.20	5.43	0.79	-0.85	1.70
25	75	75	50	0.56	1.21	0.93	3.16	3.05	2.38	4.18	3.46	5.50	0.88	-0.66	1.91
25	75	75	75	0.56	1.21	0.93	1.79	2.83	2.33	2.66	2.96	5.24	0.83	0.09	2.52
50	25	25	25	0.08	1.26	0.92	0.04	0.03	0.27	1.22	2.82	2.28	1.05	1.54	1.10
50	25	25	50	0.08	1.26	0.92	0.01	0.02	0.15	0.95	2.88	1.83	0.79	1.75	0.94
50	25	25	75	0.08	1.26	0.92	0.17	0.11	0.04	0.93	3.26	2.87	0.82	2.17	2.05
50	25	50	25	0.24	1.25	0.92	3.54	4.27	3.92	4.33	5.45	5.36	0.94	0.97	0.71
50	25	50	50	0.24	1.25	0.92	1.90	1.68	0.60	2.90	3.43	2.38	0.77	1.16	0.55
50	25	50	75	0.24	1.25	0.92	0.41	0.36	0.04	1.07	2.81	2.35	0.81	1.78	1.32
50	25	75	25	0.56	1.21	0.93	4.80	4.84	4.98	5.14	5.99	5.75	0.46	0.08	0.24
50	25	75	50	0.56	1.21	0.93	4.85	4.98	4.73	5.20	6.19	5.58	0.49	0.19	0.04
50	25	75	75	0.56	1.21	0.93	2.93	2.70	1.90	3.57	3.98	3.34	0.74	0.63	0.12
50	50	25	25	0.08	1.26	0.92	0.02	0.03	0.03	1.30	2.83	2.77	1.14	1.70	1.81
50	50	25	50	0.08	1.26	0.92	0.09	0.02	0.03	1.09	3.05	2.66	0.95	1.98	1.74
50	50	25	75	0.08	1.26	0.92	0.02	0.04	0.04	1.10	3.29	3.16	0.96	2.23	2.20
50	50	50	25	0.24	1.25	0.92	2.09	2.86	2.30	3.83	3.54	4.45	1.05	0.99	1.38
50	50	50	50	0.24	1.25	0.92	2.03	1.00	2.05	3.03	3.25	3.71	0.96	1.21	1.20
50	50	50	75	0.24	1.25	0.92	0.36	0.14	0.04	1.20	2.87	2.66	0.96	1.84	1.75
50	50	75	25	0.56	1.21	0.93	4.85	4.98	4.73	5.35	5.81	5.99	0.47	-0.22	0.39
50	50	75	50	0.56	1.21	0.93	4.85	4.99	4.74	5.33	5.83	6.07	0.49	-0.13	0.32
50	50	75	75	0.56	1.21	0.93	2.72	2.70	1.83	3.45	3.85	3.83	0.73	0.46	0.77
50	75	25	25	0.08	1.26	0.92	0.07	0.02	0.01	1.47	2.89	4.42	1.33	1.79	3.58
50	75	25	50	0.08	1.26	0.92	0.01	0.02	0.02	1.33	3.04	4.53	1.24	1.93	3.73
50	75	25	75	0.08	1.26	0.92	0.02	0.02	0.03	1.12	2.79	3.76	1.16	1.71	2.92
50	75	50	25	0.24	1.25	0.92	0.38	0.08	0.01	2.20	2.34	4.68	1.34	0.86	3.16
50	75	50	50	0.24	1.25	0.92	0.38	0.08	0.08	1.48	2.38	4.86	1.26	1.15	3.33
50	75	50	75	0.24	1.25	0.92	0.02	0.02	0.03	1.47	2.44	3.95	1.07	1.39	2.65
50	75	75	25	0.56	1.21	0.93	3.16	3.05	2.38	4.40	3.78	4.97	0.73	-0.83	1.36
50	75	75	50	0.56	1.21	0.93	3.16	3.05	2.38	4.12	3.62	5.05	0.70	-0.78	1.56
50	75	75	75	0.56	1.21	0.93	1.79	2.83	2.33	2.54	3.07	5.00	0.82	0.10	2.17
75	25	25	25	0.08	1.26	0.92	0.01	0.13	0.04	1.26	2.75	2.46	0.93	1.65	1.57
75	25	25	50	0.08	1.26	0.92	0.01	0.02	0.16	0.91	2.90	2.91	0.70	1.80	1.61
75	25	25	75	0.08	1.26	0.92	0.04	0.03	0.04	0.94	3.21	3.28	0.81	2.09	2.06
75	25	50	25	0.24	1.25	0.92	2.44	0.13	3.04	3.43	2.84	5.66	0.78	1.10	0.89
75	25	50	50	0.24	1.25	0.92	0.81	1.18	2.52	1.75	3.02	5.00	0.61	1.39	0.88
75	25	50	75	0.24	1.25	0.92	0.09	0.03	0.12	1.02	3.01	2.42	0.77	1.82	1.47
75	25	75	25	0.56	1.21	0.93	3.92	4.28	4.36	4.64	5.19	4.67	0.56	0.25	-0.56
75	25	75	50	0.56	1.21	0.93	3.92	4.28	4.36	4.79	5.27	4.59	0.57	0.34	-0.60
75	25	75	75	0.56	1.21	0.93	1.97	3.35	1.57	2.43	3.03	2.44	0.78	0.90	0.03
75	50	25	25	0.08	1.26	0.92	0.03	0.09	0.04	1.17	2.84	3.36	1.06	1.66	2.16
75	50	25	50	0.08	1.26	0.92	0.03	0.02	0.03	1.08	3.09	3.01	0.84	1.87	2.05
75	50	25	75	0.08	1.26	0.92	0.04	0.03	0.04	1.08	3.29	3.32	0.94	2.11	2.06
75	50	50	25	0.24	1.25	0.92	1.09	0.75	2.53	2.32	2.26	5.84	0.87	1.14	1.29
75	50	50	50	0.24	1.25	0.92	1.02	1.18	2.71	1.64	3.00	4.99	0.76	1.32	1.17
75	50	50	75	0.24	1.25	0.92	0.09	0.03	0.10	1.12	3.09	2.80	0.89	1.85	1.61
75	50	75	25	0.56	1.21	0.93	3.94	4.28	4.47	4.75	4.90	5.00	0.51	0.13	-0.56
75	50	75	50	0.56	1.21	0.93	3.94	4.28	4.43	4.77	4.87	4.90	0.49	0.07	-0.33
75	50	75	75	0.56	1.21	0.93	1.96	3.35	1.32	2.33	2.93	2.93	0.66	0.71	0.13
75	75	25	25	0.08	1.26	0.92	0.07	0.02	0.02	1.35	2.66	4.40	1.28	1.64	3.42
75	75	25	50	0.08	1.26	0.92	0.07	0.02	0.02	1.31	2.92	4.41	1.26	1.91	3.44
75	75	25	75	0.08	1.26	0.92	0.06	0.07	0.03	1.32	3.11	4.10	1.05	2.21	3.19
75	75	50	25	0.24	1.25	0.92	0.07	0.02	0.02	1.57	2.10	3.70	1.22	1.00	2.70
75	75	50	50	0.24	1.25	0.92	0.07	0.02	0.04	1.44	2.38	3.83	1.12	1.20	2.88
75	75	50	75	0.24	1.25	0.92	0.06	0.20	0.07	1.33	2.74	3.95	1.13	1.69	3.01
75	75	75	25	0.56	1.21	0.93	3.49	3.63	3.64	4.17	3.83	4.54	0.76	-0.47	0.32
75	75	75	50	0.56	1.21	0.93	3.49	3.71	3.64	4.05	3.96	4.72	0.67	-0.53	-0.03
75	75	75	75	0.56	1.21	0.93	1.83	2.32	1.56	2.88	3.08	4.53	0.67	0.20	0.98

Note: The values displayed are the measured values multiplied by 10

Table C.18. Marginal Cost of Major Therapeutic Procedures

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	2.64	2.57	1.97	0.42	0.62	0.88	8.19	6.40	7.64	4.65	3.20	4.62
25	25	25	50	2.96	2.91	2.18	4.36	3.82	4.03	11.70	9.06	9.69	4.43	2.51	3.72
25	25	25	75	3.82	3.86	2.92	4.90	4.99	4.97	12.74	11.18	10.30	3.85	2.30	2.57
25	25	50	25	2.64	2.57	1.97	0.01	0.01	0.05	7.14	6.10	6.42	4.39	3.29	4.38
25	25	50	50	2.96	2.91	2.18	2.42	2.39	2.71	10.12	8.21	8.60	4.26	2.73	3.66
25	25	50	75	3.82	3.86	2.92	4.99	4.98	4.97	12.87	11.18	10.60	4.00	2.26	2.80
25	25	75	25	2.64	2.57	1.97	0.02	0.20	0.03	7.14	6.18	5.61	4.39	3.36	3.47
25	25	75	50	2.96	2.91	2.18	0.34	0.39	0.27	8.13	6.31	6.75	4.34	3.15	3.47
25	25	75	75	3.82	3.86	2.92	4.07	4.03	4.75	12.34	10.32	10.57	4.24	2.79	2.85
25	50	25	25	2.64	2.57	1.97	0.01	0.02	0.04	7.25	5.63	6.18	4.53	2.82	4.03
25	50	25	50	2.96	2.91	2.18	3.27	1.78	1.64	9.89	6.89	6.85	4.34	2.18	3.39
25	50	25	75	3.82	3.86	2.92	4.99	4.99	4.97	12.64	10.79	10.24	3.78	1.92	2.43
25	50	50	25	2.64	2.57	1.97	0.09	0.05	0.08	7.09	5.93	5.92	4.37	3.00	3.81
25	50	50	50	2.96	2.91	2.18	1.31	1.16	0.28	9.88	7.44	6.72	4.20	2.41	3.33
25	50	50	75	3.82	3.86	2.92	4.99	4.98	4.97	12.72	10.75	10.31	3.87	1.91	2.55
25	50	75	25	2.64	2.57	1.97	0.02	0.08	0.01	7.18	5.98	5.06	4.39	3.17	3.16
25	50	75	50	2.96	2.91	2.18	0.03	0.11	0.18	7.38	6.08	6.26	4.26	2.90	3.28
25	50	75	75	3.82	3.86	2.92	4.07	4.03	4.60	12.08	10.02	10.61	3.99	2.34	2.89
25	75	25	25	2.64	2.57	1.97	0.01	0.01	0.02	7.09	4.85	3.88	4.30	2.08	2.00
25	75	25	50	2.96	2.91	2.18	0.40	0.01	0.06	7.25	4.65	4.14	3.96	1.50	1.97
25	75	25	75	3.82	3.86	2.92	4.62	2.77	4.22	11.73	7.82	7.30	3.83	1.34	1.35
25	75	50	25	2.64	2.57	1.97	0.05	0.01	0.05	6.92	5.04	3.91	4.11	2.36	1.89
25	75	50	50	2.96	2.91	2.18	0.05	0.06	0.09	7.18	4.66	4.07	3.92	1.58	1.69
25	75	50	75	3.82	3.86	2.92	4.62	2.88	4.22	11.81	7.74	7.13	3.69	1.36	1.17
25	75	75	25	2.64	2.57	1.97	0.03	0.02	0.02	6.84	5.23	4.03	4.06	2.57	1.98
25	75	75	50	2.96	2.91	2.18	0.03	0.02	0.02	7.12	5.06	4.07	3.94	2.01	1.95
25	75	75	75	3.82	3.86	2.92	3.90	2.33	3.98	11.92	7.93	7.80	3.71	1.70	1.62
50	25	25	25	2.64	2.57	1.97	0.07	0.01	0.01	7.44	6.04	6.89	4.69	3.17	4.32
50	25	25	50	2.96	2.91	2.18	4.50	3.97	3.42	11.71	9.16	9.62	4.56	2.56	3.48
50	25	25	75	3.82	3.86	2.92	4.84	4.99	4.97	12.71	11.38	10.34	3.83	2.49	2.64
50	25	50	25	2.64	2.57	1.97	0.02	0.01	0.02	7.20	6.12	6.16	4.43	3.28	4.10
50	25	50	50	2.96	2.91	2.18	2.85	2.64	3.17	10.07	8.28	8.63	4.35	2.80	3.47
50	25	50	75	3.82	3.86	2.92	4.99	4.95	4.97	12.86	11.37	10.49	4.00	2.50	2.73
50	25	75	25	2.64	2.57	1.97	0.02	0.21	0.07	7.17	6.19	5.64	4.35	3.37	3.39
50	25	75	50	2.96	2.91	2.18	0.22	0.34	0.43	7.52	6.50	5.87	4.36	3.27	3.40
50	25	75	75	3.82	3.86	2.92	4.07	4.03	4.75	12.39	10.55	10.48	4.26	2.83	2.89
50	50	25	25	2.64	2.57	1.97	0.05	0.03	0.04	7.34	5.76	5.91	4.60	2.86	3.75
50	50	25	50	2.96	2.91	2.18	2.79	1.78	2.02	9.86	7.18	6.53	4.46	2.31	3.18
50	50	25	75	3.82	3.86	2.92	4.99	4.97	4.97	12.62	11.01	10.20	3.77	2.15	2.43
50	50	50	25	2.64	2.57	1.97	0.09	0.04	0.10	7.16	5.89	5.67	4.43	3.06	3.55
50	50	50	50	2.96	2.91	2.18	0.95	1.73	0.34	9.07	7.30	6.12	4.28	2.56	3.14
50	50	50	75	3.82	3.86	2.92	4.99	4.94	4.97	12.69	10.99	10.33	3.82	2.13	2.57
50	50	75	25	2.64	2.57	1.97	0.03	0.08	0.03	7.15	6.09	4.93	4.41	3.28	2.97
50	50	75	50	2.96	2.91	2.18	0.03	0.11	0.07	7.38	6.21	5.50	4.30	3.06	3.08
50	50	75	75	3.82	3.86	2.92	4.40	4.03	4.60	12.11	10.24	10.61	4.01	2.48	2.99
50	75	25	25	2.64	2.57	1.97	0.01	0.01	0.02	7.13	5.02	3.84	4.24	2.20	1.92
50	75	25	50	2.96	2.91	2.18	0.40	0.01	0.06	7.32	4.90	4.00	4.04	1.74	1.26
50	75	25	75	3.82	3.86	2.92	4.62	2.77	4.00	11.59	7.98	6.99	3.85	1.63	1.31
50	75	50	25	2.64	2.57	1.97	0.05	0.01	0.02	6.95	5.24	3.87	4.19	2.22	1.79
50	75	50	50	2.96	2.91	2.18	0.05	0.04	0.09	7.27	4.88	4.00	3.96	1.73	1.63
50	75	50	75	3.82	3.86	2.92	4.62	2.77	4.00	11.71	7.97	7.10	3.75	1.56	1.06
50	75	75	25	2.64	2.57	1.97	0.03	0.02	0.02	6.89	5.66	3.94	4.07	2.56	1.94
50	75	75	50	2.96	2.91	2.18	0.03	0.02	0.02	7.09	5.31	4.08	3.94	2.12	1.84
50	75	75	75	3.82	3.86	2.92	3.90	2.33	3.98	11.60	8.25	7.70	3.69	1.92	1.45
75	25	25	25	2.64	2.57	1.97	0.09	0.01	0.06	7.64	6.17	5.24	4.79	3.33	3.02
75	25	25	50	2.96	2.91	2.18	0.60	2.17	0.80	9.34	7.64	5.75	4.56	3.16	2.80
75	25	25	75	3.82	3.86	2.92	4.95	4.37	4.97	12.74	11.01	10.33	3.94	2.94	2.22
75	25	50	25	2.64	2.57	1.97	0.02	0.01	0.03	7.39	6.27	4.97	4.54	3.46	2.98
75	25	50	50	2.96	2.91	2.18	0.97	1.24	0.16	8.65	7.05	5.65	4.43	3.33	2.79
75	25	50	75	3.82	3.86	2.92	4.94	4.37	4.97	12.89	11.24	10.16	4.15	3.03	2.40
75	25	75	25	2.64	2.57	1.97	0.01	0.01	0.07	7.09	6.50	4.90	4.35	3.67	2.82
75	25	75	50	2.96	2.91	2.18	0.01	0.01	0.07	7.45	6.86	5.19	4.34	3.68	2.88
75	25	75	75	3.82	3.86	2.92	4.93	2.60	4.71	13.18	11.58	10.30	4.38	3.48	3.33
75	50	25	25	2.64	2.57	1.97	0.07	0.02	0.10	7.84	6.08	4.78	4.70	3.17	2.65
75	50	25	50	2.96	2.91	2.18	0.23	1.66	0.28	8.35	7.20	5.42	4.42	3.01	2.53
75	50	25	75	3.82	3.86	2.92	4.95	4.37	4.98	12.63	11.08	9.93	3.83	2.75	2.20
75	50	50	25	2.64	2.57	1.97	0.12	0.03	0.03	7.45	6.26	4.66	4.50	3.42	2.57
75	50	50	50	2.96	2.91	2.18	0.18	1.24	0.17	8.07	6.92	5.38	4.34	3.17	2.46
75	50	50	75	3.82	3.86	2.92	4.94	4.37	4.98	12.70	11.27	10.18	3.99	2.75	2.47
75	50	75	25	2.64	2.57	1.97	0.04	0.01	0.04	7.18	6.42	4.50	4.36	3.65	2.44
75	50	75	50	2.96	2.91	2.18	0.04	0.01	0.05	7.43	6.73	4.77	4.28	3.59	2.61
75	50	75	75	3.82	3.86	2.92	4.84	2.60	4.88	12.88	11.49	9.95	4.18	3.36	3.12
75	75	25	25	2.64	2.57	1.97	0.02	0.05	0.02	6.99	5.49	3.48	4.24	2.65	1.38
75	75	25	50	2.96	2.91	2.18	0.02	0.10	0.02	7.16	5.48	3.60	4.10	2.28	1.31
75	75	25	75	3.82	3.86	2.92	4.71	2.83	3.09	11.53	9.43	6.69	3.52	2.37	0.91
75	75	50	25	2.64	2.57	1.97	0.02	0.05	0.02	7.02	5.72	3.36	4.19	2.86	1.46
75	75	50	50	2.96	2.91	2.18	0.02	0.10	0.03	7.19	5.64	3.47	4.01	2.38	1.45
75	75	50	75	3.82	3.86	2.92	4.71	3.00	3.20	11.55	9.66	6.71	3.74	2.26	1.06
75	75	75	25	2.64	2.57	1.97	0.02	0.03	0.02	6.73	5.99	3.73	3.98	3.14	1.71
75	75	75	50	2.96	2.91	2.18	0.02	0.05	0.02	6.96	6.11	3.91	3.88	2.95	1.50
75	75	75	75	3.82	3.86	2.92	3.41	1.62	2.86	10.66	9.38	7.00	3.68	2.89	1.60

Note: The values displayed are the measured values multiplied by 10