**Working paper**

# Power Curve Estimation: Functional Estimation Imposing the Regular Ultra Passum Law

Hoon Hwangbo, Andrew L. Johnson, Yu Ding
Department of Industrial & Systems Engineering, Texas A&M University, College Station, Texas 77840
{hhwangbo@tamu.edu, ajohnson@tamu.edu, yuding@iemail.tamu.edu}

Imposing economic relationships such as the Regular Ultra Passum (RUP) law improves the statistical efficiency of nonparametric estimators in finite samples. RUP law bears relevance in engineering applications such as power curve estimation in the wind energy industry. Unfortunately, the few estimators known to satisfy the RUP law are based on deterministic assumptions that do not allow noise in the modeling. In most engineering applications, however, data are inevitably noisy, due to equipment calibration, natural variations, or other issues. Thus, we propose an estimator that satisfies the RUP law while also capable of handling noisy data. We use Monte Carlo simulations to show that the proposed estimator outperforms existing deterministic estimators, particularly when the scale of noise is large. We use the proposed method to estimate a power curve considering approximately 13,000 observations of a wind turbine. The results demonstrate that the proposed estimator is well suited for engineering applications with a high degree of noise.

*Key words*: convex nonparametric least squares, curve fitting with shape constraints, nonparametric statistics, production economics analysis, wind energy application

*History*: This paper was first submitted on June 18, 2015.

## 1. Introduction

The Regular Ultra Passum (RUP) law, first described by Frisch (1965), is critical to modeling production functions which have increasing marginal rates of productivity followed by decreasing rates of marginal productivity. Applications for which the RUP law is relevant include those where worker specialization leads to increasing productivity and functions with engineering principles that dictate shape restrictions. In our research, we are particularly motivated by studying the power

curve of a wind turbine that describes the relationship between electricity output and wind speed. Aerodynamic properties of wind turbines and their control mechanism dictate that this functional relationship will have a convex shape followed by a concave shape, suggesting that imposing the RUP law will improve the finite sample performance when estimating a power curve.

Typical nonparametric estimators of production functions do not guarantee that the underlying structures of the production functions are preserved. Previous work on non-convex estimation such as Bogetoft et al. (2000) and Park and Simar (1994) allows for the possibility that the function estimated satisfies the RUP law, but does not impose the RUP law in the estimation procedure.

Very few estimators are known to satisfy the RUP law explicitly. The exception is the deterministic or data envelopment analysis (DEA) type estimators developed by Olesen and Petersen (2013) and Olesen and Ruggiero (2014). Olesen and Ruggiero (2014) (hereafter referred to as O&R), treating their data as noise free and imposing homotheticity on the input sets, develop a three-step procedure for producing an RUP-satisfying estimator, which entails 1) estimate a base isoquant; 2) locate the inflection point on the production function; and 3) estimate the convex and concave regions of the production function assuring that the production function is continuous and has only a single inflection point.

Our research objective is to develop an estimator that satisfies the RUP law while also capable of handling noisy data. The motivation of doing so is obvious, as in many applications, data are imperfect measures of the variables of interest creating noise in measurement. Data noise becomes unavoidable in engineering applications such as estimating power curves using noisy wind turbine data.

Our proposed estimator is an extension of O&R and will follow the three-step procedure they previously developed. This does not mean that our extension is straightforward; on the contrary, all three steps need to be altered for an estimation procedure that includes a model of noise. Our estimator also moves beyond standard stochastic frontier analysis (SFA) approaches to incorporating inefficiency, as our estimator allows for heteroscedasticity in both the inefficiency and random noise terms.

Similar to O&R, we still consider the single output case and impose homothetic input sets. Extension to multiple output settings is possible, if we impose separability between inputs and outputs; for additional details see Appendix EC.1. We also impose homotheticity on the input sets. Homothetic functions characterize a wide class of production functions in which the core function, characterized by constant returns-to-scale (CRS) (homogeneous of degree 1 in mathematical terms), is made more flexible by transforming the aggregate input (the output from the core function) using a scaling function that is monotonic. Homotheticity is one of the most common functional restrictions (Chambers and Mitchell 2001) used in productivity and efficiency analysis.

The remainder of this paper is organized as follows. Section 2 introduces the production frontier model and the assumptions in our research development. Sections 3, 4 and 5 present, respectively, the new version of base isoquant estimation, inflection point estimation, and production frontier function estimation, along with the adaptations needed to make the resulting estimator capable of modeling noise. Section 6 discusses the Monte Carlo simulation results and demonstrates the estimator's improved performance over its deterministic counterpart when modeling noisy data. Section 7 employs the resulting method to estimate a power curve for a wind turbine. Section 8 concludes the paper. All proofs and supplemental materials are given in the online Appendix EC.1–EC.3.

## 2. Model description

We specify the production frontier model under the assumption of homothetic input sets, including inefficiency and random noise affecting output levels, as:

$$y = \phi(\boldsymbol{X}) - u + v = F\left(g(\boldsymbol{X})\right) - u + v, \tag{1}$$

where $y$ is a random variable measuring output, $\boldsymbol{X}$ is an input vector consisting of $p$ inputs, and $u$ and $v$ are independently distributed random variables modeling inefficiency and random noise, respectively. We assume $u$ is a non-negative random variable, and $v$ has a symmetric distribution with a mean of 0. We refer to the expectation of $u$ as the mean inefficiency, denoted by $\mu$, i.e.,

$E(u) = \mu$. We define production function $\phi$ by the composition of two functions: a monotone increasing scaling function $F : \Re_+ \to \Re_+$ and a core function $g : \Re_+^p \to \Re_+$ which is homogenous of degree 1, i.e., $g(\gamma \boldsymbol{X}) = \gamma\, g(\boldsymbol{X})$ for some positive scalar $\gamma$. The production function $\phi$ is to be estimated by using $n$ pairs of data samples $\{(\boldsymbol{X}_1, y_1), \cdots, (\boldsymbol{X}_n, y_n)\}$.

We are interested in estimating a production function $\phi$ satisfying the RUP law. First, we define the scale elasticity of the production function, $\varepsilon(\boldsymbol{X}, y)$ as $\sum_{q=1}^{p} (\partial\phi(\boldsymbol{X})/\partial x_q) \cdot (x_q/y)$ where $x_q$ is the $q$th input. Then, we present the formal definition of the RUP law.

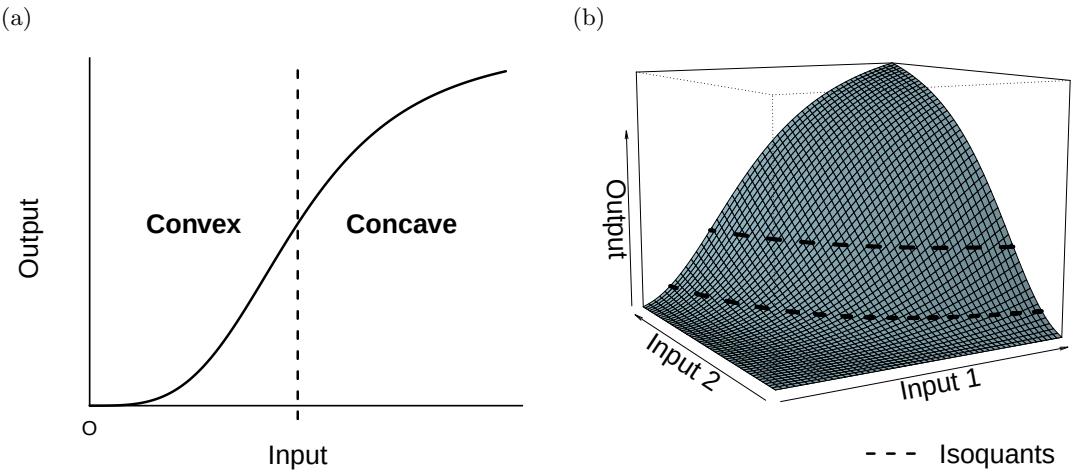
Definition 1 (Førsund and Hjalmarsson 2004). Let a single output $y$ be produced from a vector of inputs $\boldsymbol{X}$ according to a production function $\phi(\boldsymbol{X})$. The production function obeys the RUP law if $\partial\, \varepsilon(\boldsymbol{X}, y)/\partial x_q < 0$ for $\forall q = 1, \ldots, p$, and for some point $(\boldsymbol{X}_1, y_1)$ we have $\varepsilon(\boldsymbol{X}_1, y_1) > 1$, and for some point $(\boldsymbol{X}_2, y_2)$, where $\boldsymbol{X}_2 > \boldsymbol{X}_1$, $y_2 > y_1$, we have $\varepsilon(\boldsymbol{X}_2, y_2) < 1$.

Note that since $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are vectors, the inequality implies that every component of $\boldsymbol{X}_2$ is greater than or equal to every component of $\boldsymbol{X}_1$.

Production functions defined over a single (aggregate) input that satisfy the RUP law are often referred to as "S-shaped function" because of their shape. Figure 1(a) shows an example of an S-shaped 2-dimensional production function, and Figure 1(b) shows a 3-dimensional production function (i.e., two inputs) satisfying the RUP law. The power curve used in wind energy application is a typical S-shaped function like the curve shown in Figure 1(a).

We cannot apply the common methods used to estimate convex sets to estimate a function satisfying the RUP law, because the region under the production function does not form such a set. On the other hand, noticing that the S-shaped, 2-dimensional production function comprises a convex region and a concave region defining convex sets for each, we can use existing methods for each region separately and combine the two estimates. This idea, however, cannot be applied directly to the high-dimensional production functions. As illustrated in Figure 1(b), the space above the production function for low input levels is not a convex set (Frisch 1965). To circumvent this difficulty, a common treatment, as has been done by O&R, is to assume homotheticity on

**Figure 1** (Color online) Production functions satisfying the Regular Ultra Passum law




*Note.* (a) S-shaped 2-dimensional production function that satisfies the Regular Ultra Passum law and has a region where the production function is convex followed by a region where the function is concave; (b) 3-dimensional production function with two inputs and a single output where the function is concave above the isoquant containing the inflection points and the region below the isoquant containing the inflection points is neither convex nor its complement.

inputs and then aggregate the multi-dimensional inputs into a single-dimensional input, so that the high-dimensional estimation problem is reduced to a 2-dimensional problem.

To estimate a monotone convex function and a monotone concave function, we use Convex Nonparametric Least Squares (CNLS). Hildreth (1954) was the first to consider nonparametric regression subject to monotonicity and concavity constraints in the case of a single input variable $x$; see also Holloway (1979) and Hanson and Pledger (1976) for some statistical properties. Banker and Maindiratta (1992) proposed a Hildreth type function with multiple regressors using a maximum likelihood estimator. Kuosmanen (2008) proposed the least squares formulation and estimator and coined the term, Convex Nonparametric Least Squares (CNLS); see Lim and Glynn (2012) and Seijo and Sen (2011) for some statistical characteristics of CNLS; Aguilera et al. (2011) for convergence properties of CNLS; and Mammen (1991) and Mammen and Thomas-Agnan (1999), who showed that CNLS achieved the optimal nonparametric rate of convergence.

## 3. Identification and estimation of the base isoquant

We define an input isoquant Isoq$L(y)$ as follows:

$$\text{Isoq}L(y) = \{\boldsymbol{X} : \phi(\boldsymbol{X}) = y\}. \tag{2}$$

With multi-dimensional inputs, the idea behind identifying an input isoquant is to aggregate the multivariate input $\boldsymbol{X}$ associated with the corresponding univariate output $y$, as the monotonicity implies that more input produces more output. Without noise and inefficiency, there is a one-to-one mapping between the isoquant Isoq$L(y)$ and the output value $y$. Once the isoquants are identified, given the homotheticity assumption, a univariate score $\theta(\boldsymbol{X}, y)$ can be used to aggregate the multi-dimensional inputs, so that the original multivariate problem with data pairs $(\boldsymbol{X}, y)$ can be replaced by the two-dimensional problem with $(\theta, y)$ in the subsequent analysis. Under the deterministic setting, the univariate input score $\theta(\boldsymbol{X}, y)$ is calculated as follows (Olesen and Ruggiero 2014). First choose a base output level $y_0$ and identify the set of $\boldsymbol{X}$'s constructing Isoq$L(y_0)$. Then, $\theta(\boldsymbol{X}_k, y_0)$ for some $k$ is estimated by the ratio of the norm of $\boldsymbol{X}_k$ to the norm of the projection of $\boldsymbol{X}_k$ onto Isoq$L(y_0)$, i.e., $\theta(\boldsymbol{X}_k, y_0) = \|\boldsymbol{X}_k\|/\|\tilde{\boldsymbol{X}}_k\|$ where $\|\cdot\|$ denotes the 2-norm, $\tilde{\boldsymbol{X}}_k = \lambda\boldsymbol{\mathcal{I}}_1 + (1-\lambda)\boldsymbol{\mathcal{I}}_2$ for $\boldsymbol{\mathcal{I}}_1, \boldsymbol{\mathcal{I}}_2 \in \text{Isoq}L(y_0)$, $\lambda \in [0,1]$, and $\tilde{\boldsymbol{X}}_k = \gamma\boldsymbol{X}_k$ for some $\gamma \in \Re_+$.

In our problem, $y$ is stochastic, differing from the production function $\phi$ by the inefficiency and random noise terms. This raises two questions: 1) what $y$ value to use as the base output level, and 2) how to estimate the isoquant; both considering the presence of noise and inefficiency.

For the first question, to get the best estimate of the base isoquant we select $y$ to maximize the number of observations in its neighborhood. For this reason, we choose the output level that maximizes the density of the output. First, we estimate the density of $y$, $\pi(y)$, using a kernel density estimator (Rosenblatt 1956, Parzen 1962) with Gaussian kernel as

$$\hat{\pi}(y) = \frac{1}{nh}\sum_{k=1}^{n} K\left(\frac{y - y_k}{h}\right),$$

where $n$ is the number of observations and $K(\cdot)$ is the standard normal density function, and $h$ is the bandwidth parameter in a kernel estimator, controlling the smoothness of the estimated

density function. We estimate $h$ based on the plug-in method known as Silverman's rule of thumb (Silverman 1986). Then, our choice of the base output level is the mode of the estimated density function, namely, $y^* = \text{argmax}\ \hat{\pi}(y)$.

To address the second question, we do not restrict ourselves to consider Isoq$L(y^*)$ as a set of inputs that exactly produce $y^*$ as in (2). Instead, we estimate Isoq$L(y^*)$ based on a set of inputs whose corresponding output values are in the neighborhood of $y^*$. Therefore, we select a subset of data $(y_i, \boldsymbol{X}_i)_\alpha$ constructing the $(100 \cdot \alpha)\%$ level set at $y^*$ in which the set of $\boldsymbol{X}$'s is defined as follows:

$$L(y^*)_\alpha = \left\{ \boldsymbol{X}_k :\ y_l \le y_k \le y_u,\ \int_{y_l}^{y_u} \hat{\pi}(y)dy = \alpha,\ \int_{y_l}^{y^*} \hat{\pi}(y)dy = \int_{y^*}^{y_u} \hat{\pi}(y)dy,\ k = 1, \ldots, n \right\}, \qquad (3)$$

where $y_l$ and $y_u$ are the lower and the upper bounds for the subset selection, respectively, and are determined by $\alpha$ as shown in (3). The number of data pairs in the selected subset is denoted by $n_\alpha$.

The $p$th input $x_p$ (or any other input) in Isoq$L(y^*)$ can be represented by a function of other inputs and $y^*$, i.e., $x_p = \mathcal{H}(x_1, \ldots, x_{p-1}, y^*)$, and the function $\mathcal{H}$ is uniquely identified by (2). For a given $y^*$, we can simplify the function specific to $y^*$ as $x_p = \mathcal{H}_{y^*}(x_1, \ldots, x_{p-1})$ which is monotone decreasing and convex consistent with the characteristic of an isoquant. As a result, the problem of estimating Isoq$L(y^*)$ reduces to estimating the convex function $\mathcal{H}_{y^*}$.

We assume that the members of $L(y^*)_\alpha$ produce $y^*$ instead of individual $(y_i)_\alpha$'s. Then, the difference between $(y_i)_\alpha$ and $y^*$ vicariously induces errors in inputs, and such errors are common to all $x_q$'s and applied radially. Under these assumptions, our model to estimate $\mathcal{H}_{y^*}$ is:

$$x_p = \mathcal{H}_{y^*}(x_1, \ldots, x_{p-1})_\alpha \cdot \exp(w),$$

where $w$ is the random error. Assuming $E(w) = 0$, we apply CNLS minimizing the sum of squares of log-transformed errors as:

$$\min_{\boldsymbol{\beta}} \quad \sum_{i=1}^{n_\alpha} (\ln x_{ip} - \ln \hat{x}_{ip})^2$$

$$\begin{aligned}
\text{s.t.} \quad & \hat{x}_{ip} = \beta_{i0} + \sum_{q=1}^{p-1} \beta_{iq} x_{iq}, \ \forall i = 1, \ldots, n_\alpha, \\
& \beta_{i0} + \sum_{q=1}^{p-1} \beta_{iq} x_{iq} \leq \beta_{i'0} + \sum_{q=1}^{p-1} \beta_{i'q} x_{iq}, \ \forall i, i' = 1, \ldots, n_\alpha, \\
& \beta_{iq} \leq 0, \ \forall i = 1, \ldots, n_\alpha, \forall q = 1, \ldots, p-1.
\end{aligned} \tag{4}$$

$n_\alpha$ needs to be large enough to estimate the shape of the isoquant represented by the estimate $\{(x_{i1}, \ldots, x_{i,p-1}, \hat{x}_{ip}) : \ i = 1, \ldots, n_\alpha\}$. Based on our numerical experiments, we recommend adjusting $\alpha$ so that $n_\alpha$ is never smaller than $30(p-1)$.

CNLS in (4) results in a conditional mean estimate for the population $(y_i, \boldsymbol{X}_i)_\alpha$, hereafter referred to as $\text{Isoq}L(y^*)_{\text{CM}}$ where the subscript CM implies "conditional mean". If there is no inefficiency in the data, then $\text{Isoq}L(y^*)_{\text{CM}} = \text{Isoq}L(y^*)$ asymptotically. Alternatively, if there is no noise in the data, the deterministic estimator of O&R using a subset of the full population of data by imposing the condition $y_i \geq y^*$ (see Ruggiero 1996) provides an unbiased estimator.

Among the point estimates of $\text{Isoq}L(y^*)_{\text{CM}}$, we can identify a unique set of grid points, $\boldsymbol{\mathcal{I}}_j$ for $j = 1, \ldots, n_g$; the points where the unique hyperplanes from (4) intersect and the boundary points of the convex surface of $\text{Isoq}L(y^*)_{\text{CM}}$. The grid points correspond to the $\boldsymbol{\mathcal{I}}$'s in the deterministic setting but with reduced dimensionality. With this knowledge, we can use the norm ratio between $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$, as had been used previously, to calculate the univariate input score $\theta(\boldsymbol{X}, y^*)$. The optimization problem for solving for the norm ratio using the grid points, where $\boldsymbol{\mathcal{I}}_j = (\mathcal{I}_{j1}, \ldots, \mathcal{I}_{jp})$, is as follows:

$$\begin{aligned}
\min_{\gamma, \boldsymbol{s}, \boldsymbol{\lambda}} \quad & z_k = \gamma - \delta \sum_{q=1}^{p} s_q \\
\text{s.t.} \quad & \gamma x_{kq} - \sum_{j=1}^{n_g} \lambda_j \mathcal{I}_{jq} - s_q = 0, \ \forall q = 1, \ldots, p, \\
& \sum_{j=1}^{n_g} \lambda_j = 1, \\
& \lambda_j \geq 0, \ \forall j = 1, \ldots, n_g, \\
& s_q \geq 0, \ \forall q = 1, \ldots, p,
\end{aligned} \tag{5}$$

where $z_k$ is the objective function value of the optimization problem for $\boldsymbol{X}_k$, $\delta$ is a positive small number, and $s_q$ denotes a slack variable for the $q$th input. Then, using the solutions of (5), the aggregate input estimate $\hat{\theta}(\boldsymbol{X}_k, y^*) = (z_k)^{-1}$.

REMARK 1. Homotheticity implies that $\phi(\boldsymbol{X})$ is invariant to $y^*$. In other words, there exists a monotone increasing scaling function $f$ such that $\phi(\boldsymbol{X}) = F(g(\boldsymbol{X})) = f(\theta(\boldsymbol{X}, y^*))$. Such function $f$ can be defined by $f(a) := F(a \cdot \kappa(y^*))$ where $\kappa : \Re_+ \to \Re_+$ is a monotone increasing function of $y^*$ such that $g(\boldsymbol{X}) = \kappa(y^*) \cdot \theta(\boldsymbol{X}, y^*)$.

## 4. Inflection point estimation

Given aggregate input estimate $\hat{\theta}(\boldsymbol{X}, y^*)$, we estimate the scaling function $f$ specified in Remark 1 to complete the estimation of the homothetic production frontier function $\phi$. For simplicity, we refer to aggregate input $\hat{\theta}(\boldsymbol{X}_k, y^*)$ as $x_k$ for $k = 1, \ldots, n$. Alternatively, in the single input (regressor) case, we skip the input aggregation step and simply define $x_k$ for $k = 1, \ldots, n$ as the single measured input. Then, from (1) and Remark 1, we have

$$y = \phi(\boldsymbol{X}) - u + v = f(x) - u + v.$$

With subtraction and addition of mean inefficiency $\mu$ in the right hand side,

$$y = [f(x) - \mu] + [\mu - u + v] = \psi(x) + e, \tag{6}$$

where $\psi(x) := f(x) - \mu$ is called an average-practice production function and $e := \mu - u + v$ denotes composite error with $E(e) = 0$. We estimate $f$ by estimating $\psi$ and then shifting $\psi$ upwards with the estimate of the mean inefficiency $\mu$.

For univariate or aggregate input, an S-shaped function such as $\psi$ and $f$ is estimable after dividing the support of the function into the convex region and the concave region. The point at which the function changes from convex to concave is the inflection point. For the inflection point of a deterministic production frontier, O&R apply DEA and Free Disposal Hull (FDH) to find

efficient observations and choose, as the inflection point, one of the FDH efficient points which allows the maximum number of efficient observations on the frontier. In the presence of noise, however, this approach is no longer appropriate. DEA and FDH are based on models that assume no noise, thus in the presence of noise the deterministic estimators of DEA and FDH are biased; the bias increases as the noise level increases. We now present a new procedure to estimate the inflection point accounting for noise in the stochastic setting.

We define the inflection point on an S-shaped curve as the point on the curve where the second derivative of the curve is zero and denote it by $(x^*, \psi(x^*))$. Since the second derivative of the curve is near zero around the inflection point, the production function is approximately linear in the neighborhood of the inflection point. We generate a set of linear approximations in the neighborhood of the inflection point. Once the linear approximations of the production function are identified, we can estimate the inflection point as the averaged measure from the linear approximations. Below, we explain 1) how to define the neighborhood of the inflection point, and 2) how to estimate the linear approximations of the production function leading to the inflection point estimate.

### 4.1. Neighborhood of the Inflection Point

We define the neighborhood of the inflection point as an interval in which we can confidently detect a change in derivative on the production function. Since we do not know the location of the inflection point *a priori*, we will start with partitioning the whole input region into disjoint bins of equal width; a method known as binning. Subsequently, we will enlarge the bin width (so effectively reduce the number of bins) and shift back and forth the starting position of the first bin, so that we can find an interval large enough in which we estimate a single change in the sign of the derivative. Then this interval captures the inflection point and is the neighborhood we are seeking.

Suppose we partition the input region into $m$ bins, each of which has size of $h$. We define the $b$th bin by the interval of input $[r_b, r_b + h)$ for $b = 1, \ldots, m$, where $r_b$ is the starting position of the

$b$th bin and $r_{b+1} = r_b + h$ for $b = 1, \ldots, m-1$. We use $(\eta_b, \xi_b)$ pair to denote the sample average of input and output associated with the $b$th bin, respectively, namely:

$$(\eta_b, \xi_b) = \left( \frac{1}{n_b} \sum_{k=1}^{n} x_k \mathbb{1}_{[r_b,\, r_b+h)}(x_k),\ \frac{1}{n_b} \sum_{k=1}^{n} y_k \mathbb{1}_{[r_b,\, r_b+h)}(x_k) \right),\ b = 1, \ldots, m,$$

where $\mathbb{1}_A(x)$ is an indicator function, returning 1 if $x \in A$; or 0 otherwise, and $n_b = \sum_{k=1}^{n} \mathbb{1}_{[r_b,\, r_b+h)}(x_k)$ is the number of observations in the $b$th bin.

By connecting the point estimates $(\eta_b, \xi_b)$ using linear segments, we generate a piecewise linear estimate of the average-practice function $\psi$. If we can find two boundary points, $(\eta_{c^-}, \xi_{c^-})$ and $(\eta_{c^+}, \xi_{c^+})$, respectively, such that:

$$\begin{aligned} &\frac{\xi_{b+1} - \xi_b}{\eta_{b+1} - \eta_b} - \frac{\xi_b - \xi_{b-1}}{\eta_b - \eta_{b-1}} > 0,\ \forall b = 2, \ldots, c^-,\ \textit{and} \\ &\frac{\xi_{b+1} - \xi_b}{\eta_{b+1} - \eta_b} - \frac{\xi_b - \xi_{b-1}}{\eta_b - \eta_{b-1}} < 0,\ \forall b = c^+, \ldots, m-1, \end{aligned} \tag{7}$$

where $c^+ = c^- + 1$, then the interval $(\eta_{c^- - 1}, \eta_{c^+ + 1})$ asymptotically captures the inflection point, as formalized in Lemma 1 below.

LEMMA 1. *As $n_{c^-} \to \infty$ and $n_{c^+} \to \infty$, $x^*$ exists in an open interval $(\eta_{c^- - 1}, \eta_{c^+ + 1})$.*

To find $(\eta_{c^-}, \xi_{c^-})$ and $(\eta_{c^+}, \xi_{c^+})$, we start with a small initial bin size $h$ and gradually increase it by $\Delta h$ every time when certain conditions stated below are met. Another parameter to vary is the starting position of the first bin, denoted by $r_1$. We propose to draw $r_1$ values from a uniform distribution. Algorithm 1 presents the specific procedure to locate $(\eta_{c^-}, \xi_{c^-})$ and $(\eta_{c^+}, \xi_{c^+})$ by varying $h$ and $r_1$. In implementation, we choose both $h$ and $\Delta h$ to be a hundredth of the length of input space. The maximum value of $h$ is restricted by a fourth of the length of input space creating four bins, as when $h$ is greater than that, and the number of bins fewer than four, it is no longer possible for us to test if $\psi$ has an S-shape.

**Algorithm 1** (Locating $(\eta_{c^-}, \xi_{c^-})$ and $(\eta_{c^+}, \xi_{c^+})$)

**Require:** $h$: bin size, $\Delta h$: increment of $h$

$l \leftarrow 1$
**while** $l < n_{rep}$ **do**
  Draw a uniform random number $r_1^l \sim Unif\left(\min_k x_k, \min_k x_k + h\right)$
  Generate bins based on the intervals of $[r_b^l, r_b^l + h)$ for $b = 1, \ldots, m-1$, where $r_{b+1}^l = r_b^l + h$
  Obtain point estimates $(\eta_b^l, \xi_b^l)$ for each bin, and find a pair of boundary point candidates $\left(\eta_{c^-}^l, \xi_{c^-}^l\right)$ and $\left(\eta_{c^+}^l, \xi_{c^+}^l\right)$ which satisfy (7)
  **if** Such candidates do not exist and $h$ has reached its maximum value **then**
    **break**
  **else if** Such candidates do not exist and $h$ has not reached its maximum value **then**
    $h \leftarrow h + \Delta h$
    $l \leftarrow 1$
  **else**
    $l \leftarrow l + 1$
  **end if**
**end while**
**return** $\left(\eta_{c^-}^l, \xi_{c^-}^l\right)$ and $\left(\eta_{c^+}^l, \xi_{c^+}^l\right)$ for $\forall l = 1, \ldots, n_{rep}$

Since we draw $n_{rep}$ random samples for $r_1$, this consequently produces $n_{rep}$ pairs of $(\eta_{c^-}, \xi_{c^-})$ as well as $n_{rep}$ pairs of $(\eta_{c^+}, \xi_{c^+})$. Let $\mathcal{C}_x^- = \left\{\eta_{c^-}^l : \ l = 1, \ldots, n_{rep}\right\}$ and $\mathcal{C}_x^+ = \left\{\eta_{c^+}^l : \ l = 1, \ldots, n_{rep}\right\}$ be the collection of all $\eta_{c^-}$'s and $\eta_{c^+}$'s located, and let $\mathcal{C}_x = \mathcal{C}_x^- \cup \mathcal{C}_x^+$. We can have the following property.

THEOREM 1. *Suppose that $x_k$'s are uniformly distributed in $\left[\inf_l r_{c^- - 1}^l, \sup_l r_{c^+ + 1}^l + h\right)$. Then, as $n_b \to \infty$ and $n_{rep} \to \infty$, $x^*$ exists in $(\inf \mathcal{C}_x, \sup \mathcal{C}_x)$, and $\psi$ satisfies the RUP law on the support $[\inf \mathcal{C}_x, \sup \mathcal{C}_x]$.*

Theorem 1 effectively says that $[\inf \mathcal{C}_x, \sup \mathcal{C}_x]$ is the neighborhood we try to identify around the inflection point. We want to note that the conditions in Theorem 1 are not restrictive, as these conditions just require us to have large enough samples around the inflection point and large enough replication resulting from the random draws of $r_1$.

### 4.2. Linear Approximations of a Production Function in the Neighborhood of the Inflection Point

Once the neighborhood is identified and if $\psi$ can be well approximated within the neighborhood by a linear function, then, the position of the inflection point can be reasonably estimated by the middle point on the linear approximation. Theorem 1, however, does not guarantee that $\psi$ can be approximated by a linear function within the entire support of the neighborhood. So our approach is to contract, respectively, from each boundary of the neighborhood forming a subinterval of $[\inf \mathcal{C}_x, \sup \mathcal{C}_x]$, over which linear approximation can be justified.

We consider a set of boundary point candidates resulting from Algorithm 1 which by construction are random samples from $\psi$ having an S-shape on the interval $[\inf \mathcal{C}_x, \sup \mathcal{C}_x]$. Define $\mathcal{C} = \left\{(\eta^l_{c^-}, \xi^l_{c^-}) : \ l = 1, \ldots, n_{rep}\right\} \cup \left\{(\eta^l_{c^+}, \xi^l_{c^+}) : \ l = 1, \ldots, n_{rep}\right\}$, so that the data pair summarizing the boundary point candidates can be expressed as $(\rho_t, \tau_t)$ for $t = 1, \ldots, d$, where $(\rho_t, \tau_t) \in \mathcal{C}$ for $\forall t$ and $d = 2n_{rep}$ is the cardinality of $\mathcal{C}$. Using these candidates, we fit a convex curve and a concave curve separately over the whole neighborhood. To fit the convex and concave curves, we use the CNLS estimators, as follows:

$$
\begin{aligned}
\min_{\boldsymbol{\beta}^{cvx}} \quad & \sum_{t=1}^{d} (\tau_t - \hat{\tau}_t^{cvx})^2 \\
\text{s.t.} \quad & \hat{\tau}_t^{cvx} = \beta_{t,0}^{cvx} + \beta_{t,1}^{cvx} \rho_t, \ \forall t = 1, \ldots, d, \\
& \beta_{t,0}^{cvx} + \beta_{t,1}^{cvx} \rho_t \geq \beta_{t',0}^{cvx} + \beta_{t',1}^{cvx} \rho_t, \ \forall t, t' = 1, \ldots, d, \\
& \beta_{t,1}^{cvx} \geq 0, \ \forall t = 1, \ldots, d,
\end{aligned} \tag{8}
$$

$$\begin{aligned}
\min_{\boldsymbol{\beta}^{ccv}} \quad & \sum_{t=1}^{d} \left(\tau_t - \hat{\tau}_t^{ccv}\right)^2 \\
\text{s.t.} \quad & \hat{\tau}_t^{ccv} = \beta_{t,0}^{ccv} + \beta_{t,1}^{ccv}\rho_t, \ \forall t = 1, \ldots, d, \\
& \beta_{t,0}^{ccv} + \beta_{t,1}^{ccv}\rho_t \leq \beta_{t',0}^{ccv} + \beta_{t',1}^{ccv}\rho_t, \ \forall t, t' = 1, \ldots, d, \\
& \beta_{t,1}^{ccv} \geq 0, \ \forall t = 1, \ldots, d.
\end{aligned} \tag{9}$$

The optimization problem in (8) is solved using the Afriat inequality (Afriat 1967) and it fits a monotone increasing and convex curve, while the optimization problem in (9) is solved by changing the direction of Afriat inequalities in its second set of constraints, and it fits a monotone increasing and concave curve.

Once we fit the convex and concave curves, we also fit a linear function over the same support, $[\inf \mathcal{C}_x, \sup \mathcal{C}_x]$. We contract the right boundary point to the position where the slope of the concave curve is no longer smaller than that of the fitted line. We also contract the left boundary point to the position where the slope of the convex curve is no longer smaller than that of the fitted line. We repeat the above curve fitting and boundary retreating procedure iteratively until either the right boundary becomes smaller than the left boundary or the number of observations in the subinterval is fewer than a preset threshold.

Algorithm 2 presents the detailed steps for estimating the inflection point. While implementing the algorithm, $\tilde{\mathcal{C}} = \varnothing$ occurs when the right boundary is smaller than the left boundary. The number of observations required in the subinterval, denoted by $K$, is chosen to be 30.

**Algorithm 2** (Estimating the inflection point from the set of candidates)

**Require:** $K$: preset threshold, $d$: cardinality of $\mathcal{C}$

1: **repeat**

2: Apply CNLS to fit a convex function (8) and a concave function (9) using the points in $\mathcal{C}$

3: Fit a linear function of $\tau_t = \beta_0^{lin} + \beta_1^{lin}\rho_t + \epsilon_t$ for $t = 1, \ldots, d$ using ordinary least squares

4: Find a subset $\tilde{\mathcal{C}}$ of $\mathcal{C}$ such that $\tilde{\mathcal{C}} = \left\{(\rho_t, \tau_t) : \ \hat{\beta}_{t,1}^{cvx} \geq \hat{\beta}_1^{lin}, \ \hat{\beta}_{t,1}^{ccv} \geq \hat{\beta}_1^{lin}, \ t = 1, \ldots, d\right\}$

5: Set $d$ to be the cardinality of $\tilde{\mathcal{C}}$

6: **if** $d > K$ **then**

7: $\mathcal{C} \leftarrow \tilde{\mathcal{C}}$

8: **end if**

9: **until** $d \leq K$

10: Estimate the inflection point $(\hat{x}^*, \hat{\psi}(\hat{x}^*))$ as the mean values of the $\rho_t$'s and $\tau_t$'s in $\mathcal{C}$, i.e.

$$(\hat{x}^*, \hat{\psi}(\hat{x}^*)) = \left(\sum_{t=1}^{d} \rho_t/d, \sum_{t=1}^{d} \tau_t/d\right)$$

11: **return** $(\hat{x}^*, \hat{\psi}(\hat{x}^*))$

## 5. Estimation of a production function consistent with the regular ultra-passum law

Given the inflection point estimate $(\hat{x}^*, \hat{\psi}(\hat{x}^*))$ resulting from Algorithm 2, we divide the data set into two parts: $\{(x_k, y_k) : x_k < x^*\}$ defining the convex region and $\{(x_k, y_k) : x_k > x^*\}$ defining the concave region. Simply applying CNLS to each region and integrating the estimates from both regions does not ensure continuity of the production function estimate at the inflection point. Thus, we explicitly impose the inflection point estimate to be a point on the production function and solve the CNLS problems for both regions in a unified optimization problem. In what follows, we present a method to estimate the average-practice curve in Section 5.1, to estimate the mean inefficiency in Section 5.2, and to estimate the frontier curve in Section 5.3.

### 5.1. Average-Practice Curve Estimation Using CNLS

We add the inflection point estimate $(\hat{x}^*, \hat{\psi}(\hat{x}^*))$ to the observed data set. As a result, this augmented set is expressed as $\{(x_{k'}, y_{k'}) : k' = 1, \ldots, n+1\} := \{(x_k, y_k) : k = 1, \ldots, n\} \cup (\hat{x}^*, \hat{\psi}(\hat{x}^*))$. We also rearrange the data set in non-decreasing order of the input values, and index the inflection point estimate by $k^*$, so that the observations in the convex region are indexed as from 1 to $k^* - 1$ and those in the concave region from $k^* + 1$ to $n + 1$. Using the augmented data set, we solve the extended CNLS problem as:

$$\min_{\boldsymbol{\beta}} \quad \sum_{k'=1}^{n+1} \left(y_{k'} - \hat{y}_{k'}\right)^2 \tag{10a}$$

$$\text{s.t.} \quad \hat{y}_{k'} = \beta_{k',0} + \beta_{k',1} x_{k'}, \ \forall k' = 1, \ldots, n+1, \tag{10b}$$

$$\beta_{k',1} = \frac{\hat{y}_{k'+1} - \hat{y}_{k'}}{x_{k'+1} - x_{k'}}, \ \forall k' = 1, \ldots, k^* - 1, \tag{10c}$$

$$\beta_{k',1} = \frac{\hat{y}_{k'} - \hat{y}_{k'-1}}{x_{k'} - x_{k'-1}}, \ \forall k' = k^* + 1, \ldots, n+1, \tag{10d}$$

$$\beta_{k',1} \leq \beta_{k'+1,1}, \ \forall k' = 1, \ldots, k^* - 1, \tag{10e}$$

$$\beta_{k'-1,1} \geq \beta_{k',1}, \ \forall k' = k^* + 1, \ldots, n+1, \tag{10f}$$

$$\beta_{k',1} \geq 0, \ \forall k' = 1, \ldots, n+1, \tag{10g}$$

$$\hat{y}_{k^*} = y_{k^*}, \tag{10h}$$

$$\beta_{k',0} \leq 0, \ \forall k' = 1, \ldots, k^*, \tag{10i}$$

$$\hat{y}_{k'} \geq 0, \ \forall k' = 1, \ldots, n+1. \tag{10j}$$

The constraints in (10c) – (10f), equivalent to the constraints used by Hildreth (1954), makes use of the data ordering information to improve computational efficiency (see Lee et al. 2013 for the details). In fact, the constraints in (10b), (10c), (10e), and (10g) are equivalent to those of CNLS fitting a monotone increasing and convex function with a single input, whereas (10b), (10d), (10f), and (10g) alternate the constraints of CNLS modeling for the concave part. Constraint (10h) forces the production function estimate from both regions to share the same $\beta_{k^*,1}$ and subsequently, the same $\beta_{k^*,0}$ as well. Note, having (10h) also ensures that the inflection point estimate is on the estimated production function and that the production function estimate is continuous.

Meanwhile, production functions must satisfy weak essentiality (Chambers 1988), meaning $f(\mathbf{0}) = 0$, where $\mathbf{0}$ is the null vector of inputs of proper dimensions. To be consistent with this requirement, we postulate that our production function goes through the origin. It can be proven that including (10i) and (10j) ensures the production function estimate goes through the origin (see Proposition EC.1 in Appendix EC.3 for the statement of sufficiency and proof).

### 5.2. Mean inefficiency Estimation

Before we set forth to estimate the mean inefficiency term, we want to explicitly allow heteroscedasticity in the composite error term. To do that, we write the mean inefficiency as $\mu_k$ instead of a constant $\mu$, so that the model in (6) is expressed as,

$$y_k = [f(x_k) - \mu_k] + [\mu_k - u_k + v_k] = \psi(x_k) + e_k,\ k = 1, \ldots, n, \tag{11}$$

and $f(x_k) = \psi(x_k) + \mu_k$.

The reason to allow heteroscedasticity in error modeling is the following. Imposing the RUP law implies the production function starting with a convex region, so that the production function may not deviate much from zero over the low value region of $x$. What this implies for the variance of the output is that the variance at places close to the origin is likely small, but this small level of variance is not constant throughout the production function, as the observations show larger variations at large output levels.

To model the heteroscedasticity of the composite error, we use a cluster strategy. The strategy starts with binning the input data into a number of disjoint intervals. Each input is paired with an output data point, and naturally, for all the inputs falling in the same bin, the corresponding input-output data pairs form a cluster. Subsequently, we assume that the errors are homoscedastic or identically distributed within a cluster. In implementation, we define clusters with a uniform number of observations, $n_f$. Understandably, the last bin may include less than $n_f$ observations. If there are fewer than 4 observations in the last bin, we merge the last bin into the nearest preceding bin.

The choice of $n_f$ is such that it creates as many clusters as possible but should maintain enough number of observations in each cluster. A good choice, in our experience, is to base $n_f$ on the result of the average-practice function estimation. Specifically, we let $n_f$ be the largest integer such that $n_f \leq n/m_f$, where $m_f$ is the number of hyperplanes in the average-practice function estimate. If $n/m_f$ is less than the minimum threshold (set as 4 observations per cluster), we then set $n_f$ to the minimum threshold value.

If we index the clusters by $b$ and the observations in each cluster by $o$, from (11), we have

$$e_{b,o} = (\mu_b - u_{b,o}) + v_{b,o},\ o = 1, \ldots, n_f,\ b = 1, \ldots, m_f.$$

Note that $u_{b,o}$ and $v_{b,o}$ are i.i.d. within the $b$th cluster. Since $u_{b,o}$ a non-negative random variable, the density of $(\mu_b - u_{b,o})$ has jump discontinuity at $\mu_b$, and the density is zero if $(\mu_b - u_{b,o})$ is greater than $\mu_b$. With this variable specification, we apply the Hall and Simar (2002) producer to estimate the cluster-wise mean inefficiency $\mu_b$. Hall and Simar (2002) indicate that their estimator is biased and that the bias diminishes as the scale of noise tends towards 0. Through numerical experiments they find this bias is small. Our experiments indicate similar results. According to Hall and Simar (2002), if the density of $e_{b,o}$ has jump discontinuity, the first derivative of the density of $e_{b,o}$ achieves its greatest absolute value at the point where the jump discontinuity occurs. We assume that the jump discontinuity of the density of $e_{b,o}$ is induced by that of $(\mu_b - u_{b,o})$ when the variance of $v_{b,o}$ is quite small. Then we estimate $\mu_b$ by finding a point where the first derivative of the density of $e_{b,o}$ is maximized in absolute terms. Algorithm 3 explains how to apply Hall and Simar's method for each cluster.

**Algorithm 3** (Estimating cluster-wise mean inefficiency)

**Require:** $h_{HS}$: bandwidth parameter, $\Delta h_{HS}$: decrement of $h_{HS}$, $T$: stopping criterion

1: $\hat{\mu}_b \leftarrow \max_o \hat{e}_{b,o}$
2: **repeat**
3: Estimate the density of the residuals $\hat{e}_{b,o}$ for $\forall o = 1, \ldots, n_f$ applying kernel density estimation with bandwidth $h_{HS}$.
4: Compute $\tilde{\mu}_b$ by finding a point where $\hat{p}''(e)$ is zero in a neighborhood of $\hat{\mu}_b$ where $\hat{p}(e)$ is the density estimate of the residuals
5: **if** $|\tilde{\mu}_b - \hat{\mu}_b| \leq T$ **then**
6: $h_{HS} \leftarrow h_{HS} - \Delta h_{HS}$
7: $T \leftarrow |\tilde{\mu}_b - \hat{\mu}_b|$
8: $\hat{\mu}_b \leftarrow \tilde{\mu}_b$

9: **end if**

10: **until** $|\tilde{\mu}_b - \hat{\mu}_b| > T$

11: **return** $\hat{\mu}_b$

We set the initial value of the bandwidth parameter $h_{HS}$, introduced in Hall and Simar (2002), to $(\max \hat{e}_k - \min \hat{e}_k)/3$. For Gaussian kernels, our choice implies that we assume the minimum residual is $3\sigma$ away from the maximum residual. The initial stopping criterion $T$ can be any number greater than the maximum of $|e_{b,o}|$ for $\forall b$ and $\forall o$, which is typically bounded by $\max y_k - \min y_k$. For our experiments, we set $\Delta h_{HS}$ to one hundredth of the initial value of $h_{HS}$ and use $T = 10,000$. The iterative search stops when there is a significant change on the shape of the density estimate.

### 5.3. Production Frontier Curve Estimation Using CNLS Smoothing

Using the relationship $f(x_k) = \psi(x_k) + \mu_k$, we construct a frontier function estimate, $\zeta$, employing cluster-wise mean inefficiencies, such as $\zeta(x_k) = \psi(x_k) + \sum_{b=1}^{m_f} \mu_b \mathbb{1}_{I_b}(x_k)$, where $I_b$ is a set of input values in the $b$th cluster. The values of $\zeta$, shifted upward from $\psi$ with different constants, no longer satisfy the RUP law and the shape constraints. Therefore, we need to refit the shifted cluster-wise production function segments $\zeta$ by using an extended CNLS in (10) so as to obtain a continuous frontier function estimate satisfying the shape constraints.

To use the extended CNLS in (10), one needs to know which point to use as the inflection point on the frontier curve. Under the assumption of heteroscedastic errors, the shape of $f$ differs from that of $\psi$, and the location of the inflection point on $f$ differs from that on $\psi$, too.

Our treatment here is to find the inflection point among the data pairs $(x_k, \zeta(x_k))$. Specially, we identify a point that minimizes the sum of squared errors objective function in (10). But we only want to consider the mid-points of the clusters as our candidate points, namely $(\{r_b + r_{b+1}\}/2, \zeta(\{r_b + r_{b+1}\}/2))$ for $b = 2, \ldots, m_f - 2$. Had we picked the boundary point of a cluster as the inflection point, the extended CNLS in (10) cannot guarantee that the resulting function is smooth at the inflection point. By contrast, this problem will not occur, if we use the mid-point

of a cluster as the candidates for the inflection point, as the smoothness of this point was ensured when the cluster-wise $\zeta$'s were initially estimated. Given that the number of mid-points of the clusters is limited, it is then easy to solve (10). In fact, one can simply perform exhaustive search of all inflection point candidates and pick the one resulting in the smallest sum of squared errors.

## 6. Monte Carlo simulation results

In this section, we use Monte Carlo simulations to compare the performance of our proposed estimator, referred to as HJD hereafter, to O&R's estimator based on the data generation process (DGP) proposed in O&R and its variants.

The O&R DGP generates 100 observations with two inputs and one output. For the two-dimensional input, the O&R DGP 100 draws random samples of modulus $\omega \sim Unif(0, 2.5)$ and angles $\eta \sim Unif(0.05, \pi/2 - 0.05)$ and converts them to the Cartesian coordinates, $\boldsymbol{X} = (x_1, x_2) = (\omega \cos\eta, \omega \sin\eta)$. It then calculates the output values before perturbation by inefficiency and noise according to $y = F\left(g(x_1, x_2)\right)$, where $g(x_1, x_2) = \left(\beta x_1^{(\sigma-1)/\sigma} + (1-\beta)x_2^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}$ with $\beta = 0.45$ and $\sigma = 1.51$, and $F(z) = 15/(1 + \exp(-5\ln z))$. Finally, the O&R DGP draws 100 samples of inefficiency $u$ from $N^+(0, \tilde{\sigma}_u^2)$ and applies it to input vectors, i.e., $\boldsymbol{X}' = \exp(u) \cdot \boldsymbol{X}$. Note that the actual variance of $u$ is $\sigma_u^2 = \tilde{\sigma}_u^2 (1 - 2/\pi)$. We use data $(\boldsymbol{X}', y)$ to estimate $\phi(\boldsymbol{X})$.

For our first set of experiments, we use O&R's deterministic DGP. Since noise is present in wind turbine operations and the scale of the noise is often unknown, we compare the two estimators for scenarios where the standard deviation of random noise $\sigma_v$ varies in magnitude. Specifically, for a given $\tilde{\sigma}_u$, we consider five different $\sigma_v$'s: $0.2\tilde{\sigma}_u$, $0.4\tilde{\sigma}_u$, $0.6\tilde{\sigma}_u$, $0.8\tilde{\sigma}_u$, and $1\tilde{\sigma}_u$, corresponding to the signal-to-noise ratios $\sigma_u/\sigma_v$ of 8.29, 4.15, 2.76, 2.07, and 1.66, respectively. For these cases, unlike in O&R, we include the random noise term $v$, where $v \sim N(0, \sigma_v^2)$, and apply the noise to input as $\boldsymbol{X}' = \exp(u - v) \cdot \boldsymbol{X}$. We also consider three different levels of $\tilde{\sigma}_u$, 0.05, 0.1, and 0.15 which allows us to explore how different scales of the composite error affect the performance of the two estimators. In total, we compare 18 scenarios, including the scenario where $\sigma_v = 0$.

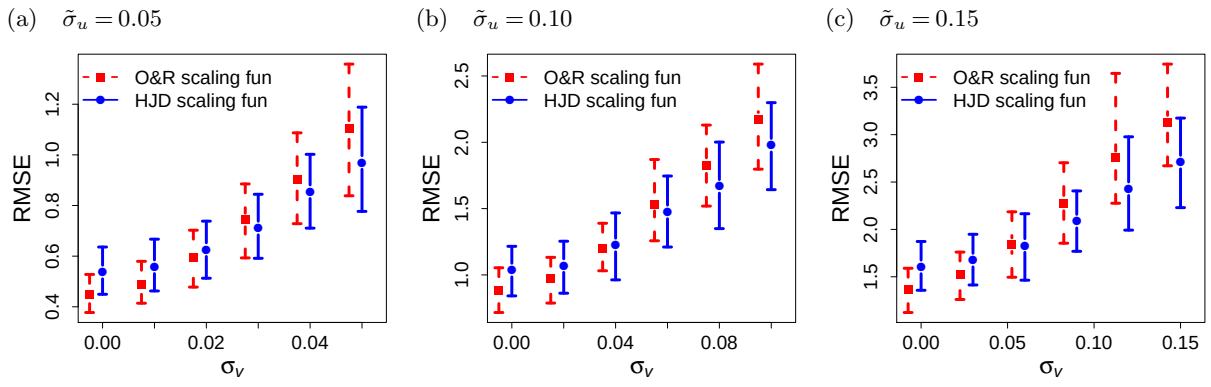
We define two stages: aggregate input estimation and S-shaped scaling function estimation including the inflection point estimation. HJD and O&R differ in both stages. We compare the two estimators in both stages as well as their overall performance. To measure the performance, we use the root mean squared error (RMSE),

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\phi(\boldsymbol{X}_k) - \hat{\phi}(\boldsymbol{X}_k)\right)^2}.$$

In the O&R DGP, the mode of output density estimate, $y^*$, is often close to zero. For small sample sizes, such as 100 observations, $\alpha$ is relatively large ($\alpha$=0.3) because we impose a minimum on the number of observations used to estimate the base isoquant. This can lead to $y_l$ being less than zero. In fact the O&R DGP generations many observations with small output levels, but large input values. Using this data leads to poor estimates of the shape of the base isoquant. To improve the estimation, we set $y_l = y^*/2$ and determine $y_u$ such that $\int_{y_l}^{y_u}\hat{\pi}(y)dy = \alpha$, when $y_l$ is calculated to be less than zero. However, asymmetrical selection of the subset does not significantly influence frontier function estimate because when estimating the scaling function the output level of the isoquant is assigned to improve the functional fit. In our experiments, a negative lower bound, $y_l$, occurs more frequently in the experiments with fewer than 350 observations.

Figure 2 demonstrates 90% credential intervals of RMSE for the second stage scaling function estimators. For this comparison, we estimate the scaling function over true univariate input $g(\boldsymbol{X})$ to avoid any bias introduced by the first stage estimator. As expected, O&R, a deterministic

**Figure 2** (Color online) Result of scaling function estimation based on true univariate input.



estimator, performs well when the scale of noise is small, or equivalently, when the signal-to-noise ratio is quite large. However, when the signal-to-noise ratio is less than or equal to 4.15 ($\sigma_v = 0.02$ for $\tilde{\sigma}_u = 0.05$), HJD becomes competitive. As the scale of composite error keeps increasing, O&R's performance deteriorates quickly.

Figure 3 compares O&R's aggregate input estimator, the true univariate input $g(\boldsymbol{X}_k)$'s, and HJD's aggregate input estimator, all applied with HJD's scaling function estimator used in the second stage. Evidently, HJD's aggregate input estimator performs comparably with, but slightly better than, O&R's aggregate input estimator.

When an aggregate input estimator is followed with using O&R's scaling function estimator in the second stage, however, using HJD's aggregate input estimator performs significantly better than

**Figure 3** (Color online) Comparison of the aggregate input estimations using HJD's scaling function estimator in the second stage.

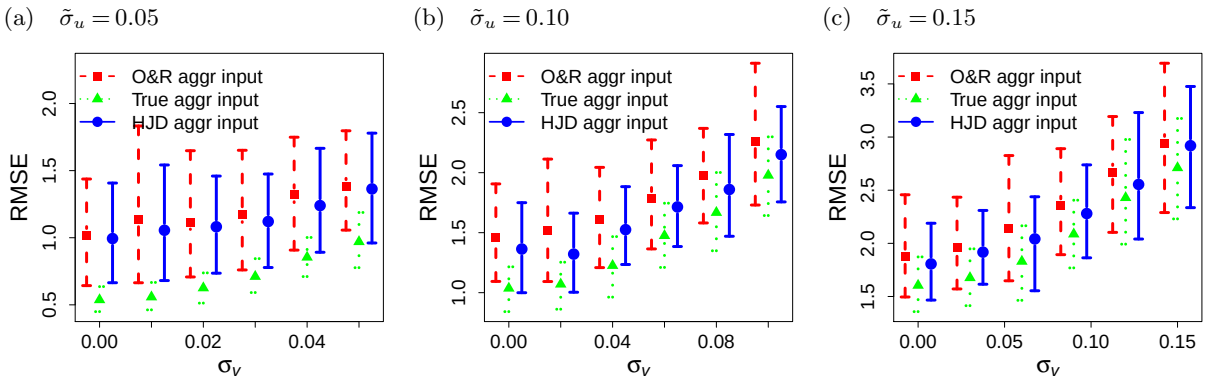
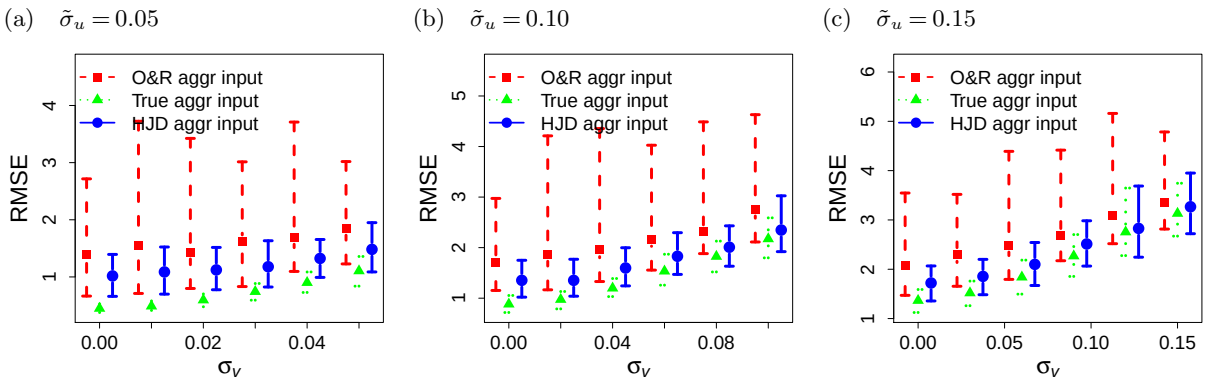


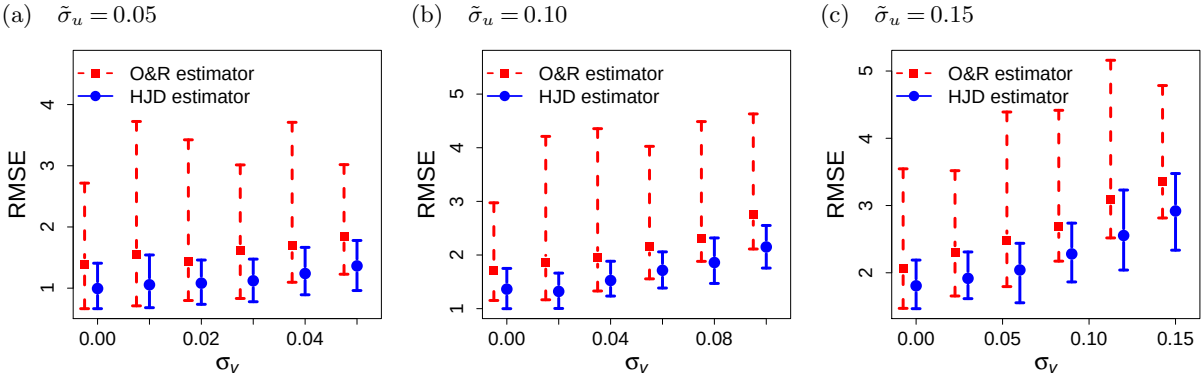

**Figure 4** (Color online) Comparison of the aggregate input estimations using O&R's scaling function estimator in the second stage.



using O&R's aggregate input; this is shown in Figure 4. This comparison outcome suggests that using HJD's scaling function estimator appears to account for errors introduced in the aggregation input estimation. From Figure 3 and 4, we also notice that the credential intervals of HJD's aggregate input estimator are comparable to those of the true univariate input; both of them are noticeably smaller than the credential intervals of O&R's aggregate input estimator, illustrating HJD's robustness.

Finally, Figure 5 presents the comparison results of the overall HJD estimator (both stage HJD) to the O&R estimator (both stage O&R), demonstrating the superior performance of HJD over O&R, when the noise level is not zero. As the signal-to-noise ratio decreases, the advantage of HJD becomes more and more pronounced.

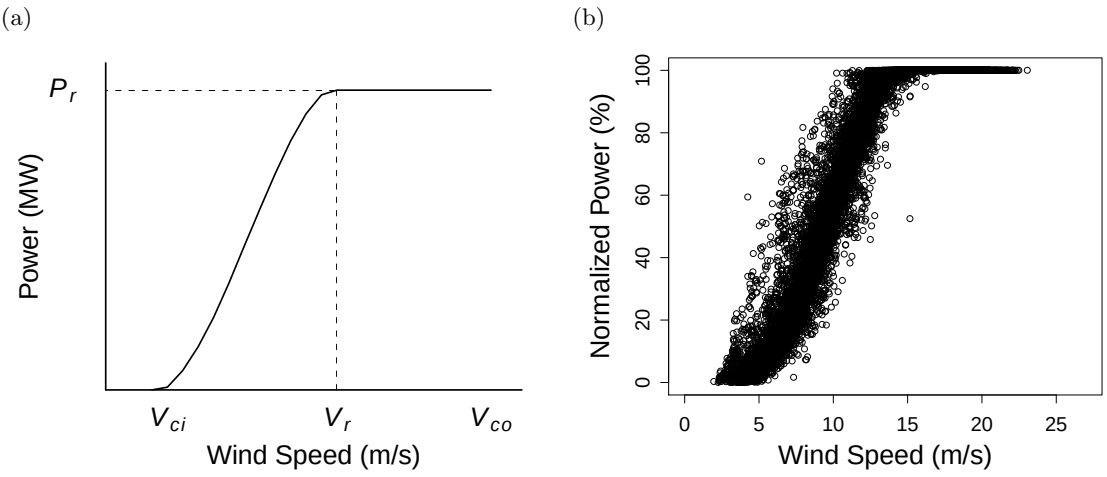
**Figure 5** (Color online) Comparison of O&R and HJD production function estimators.




## 7. Application to the estimation of the power curve for a wind turbine

We apply the proposed estimator to the estimation of the power curve of a single wind turbine. The power curve is a classic example of a production function that satisfies the RUP law. Figure 6(a) shows a typical nominal power curve provided by the turbine's manufacturer, which characterizes a turbine's power output responding to wind speeds. Starting with the cut-in wind speed $V_{ci}$, the blades start rotate and the turbine begins generating electricity. As the wind speed increases, the rates of marginal power productivity increase up to some point and then start decreasing, due to the

**Figure 6** Wind turbine data: (a) nominal power curve and (b) scatter plot of the data used in this example.




use of pitch control mechanism (which turns the turbine blades to reduce the energy absorption). Eventually, the power production levels off at the nominal power capacity of the turbine, $P_r$, known as the rated power. Equivalently, the rate of marginal power productivity becomes zero beyond the rated power. The corresponding wind speed is known as the rate speed and denoted by $V_r$. Energy production is halted when the wind speed reaches the cut-out speed $V_{co}$ for safety reasons.

While wind speed is widely accepted as the predominant factor affecting the power production of a turbine, other factors such as air density, humidity, turbulence or dusting affect the power output as well. When plotting on the wind speed-versus-power coordinates, one can observe that the actual measurements scatter broadly around the nominal curve; see Figure 6(b). Much of the randomness is attributed to the other factors that are not accounted for as well as unknown factors that also affect a turbine's power production.

The power curves are commonly used to characterize a turbine's power production performance. Upon the availability of the wind speed and power output, one naturally wonders how much systematic inefficiency there is in addition to random noise. Fitting the model in (6) can provide insights.

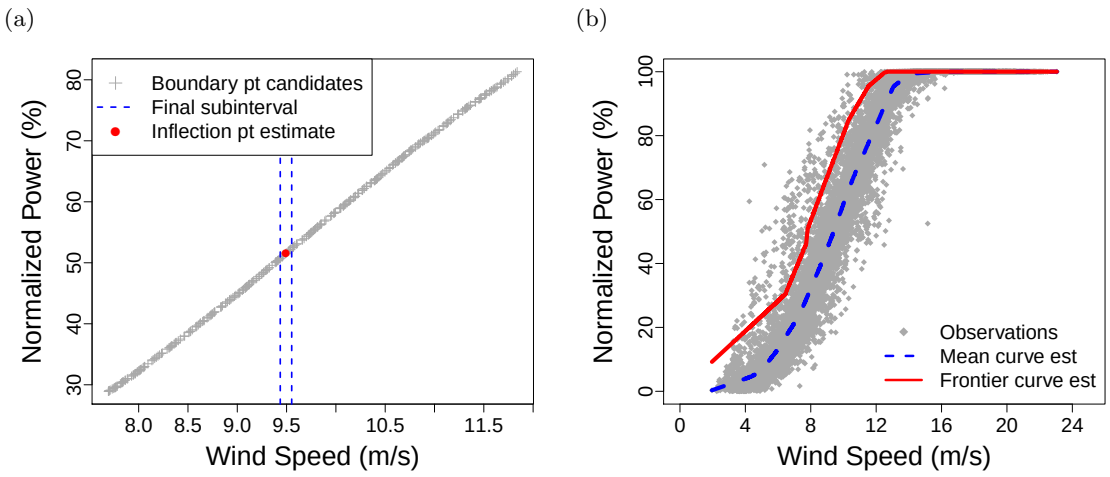
In this study, we use the data from one offshore wind turbine. The cut-in and cut-out speeds are 3.5m/s and 25m/s, respectively. The turbine has a rated wind speed around 15 m/s and the corresponding rated power is in the magnitude of megawatts, but the specific quantity cannot be

disclosed due to a confidentiality agreement. In Figure 6 and 7, we normalize the power response (vertical axis) by setting the rated power to be 100%.

We have a one year worth of the wind speed and power data, measured in 2007. In the wind industry standard practice, wind speed and power data used are the averages of 10-minute observation intervals. We have approximately 13,000 observations of both wind speed and power output after preprocessing (eliminating missing data, etc.). Even though we do not have observations for every 10-minute time interval in the year, the missing values in our sample seem to randomly occur due to measurement issues, so we do not expect any sample selection biases.

The power curve is an example of a univariate production function with only one input, namely $x$ = wind speed and $y$ = power output, so we do not need to aggregate inputs. Figure 7(a) shows the results of our inflection point estimation. With a large number of observations, the boundary point candidates form a smooth curve. In addition, the wind turbine data exhibits strong linearity in a wide range of input around the inflection point. The inflection point estimate is located almost at the center after discarding the boundary point candidates indicating slight curvatures near both

**Figure 7** Power curve estimation: (a) set of boundary point candidates and inflection point estimate, (b) production function estimates.




*Note.* No real data are observed in the range of wind speed from 0 m/s to approximately 2 m/s. Note that both the average-practice function estimate and frontier function estimate extend to (0,0) in this range.

ends of the support. Figure 7(b) shows the estimation of an average-practice production function and the frontier production function.

The average-practice and frontier function estimates can be used to evaluate turbine performance. For example, efficiency can be measured by the ratio of the area under the average-practice function estimate to the area under the production frontier estimate. In this particular example, the efficiency measure is 0.882, which implies 11.8% of systematic inefficiency on average.

Since it is difficult to estimate $\tilde{\sigma_u}$ and $\tilde{\sigma_v}$ directly from the wind turbine data without making distribution assumptions on them, we next simulate a set of data similar to the wind turbine data. We revise O&R's DGP that is used in Section 6 in order to allow heteroscedasticity in the simulated data. This way, the variance pattern of the residuals of the simulated data is consistent with that of the actual wind turbine data. The detailed procedure of the revised DGP is presented by Procedure 1.

**Procedure 1** (Revised Data Generation Process)

1: Obtain $\boldsymbol{X}$ and $\phi(\boldsymbol{X}_k)$ for $k = 1, \ldots, 100$ similar to O&R's DGP, but now with $\omega \sim Unif(0,4)$ and $F(z) = 15/(1+\exp(-3.5\ln z))$.

2: Assume that $\tilde{\sigma}_u$ and $\tilde{\sigma}_v$ are given. Sample inefficiency $u_k$ from truncated normal with mean 0, variance $\tilde{\sigma}_u^2$, minimum of 0, and maximum of $\phi(x_{1k}, x_{2k})$. For noise $v_k$, use random samples from $N(0, \sigma_{vk}^2)$, where $\sigma_{vk}$ varies depending on $g(x_{1k}, x_{2k})$. If $g(x_{1k}, x_{2k}) < 0.35$, $\sigma_{vk} = \tilde{\sigma}_v \cdot F\left(g(x_{1k}, x_{2k})\right)/F(0.35)$. On the other hand, if $g(x_{1k}, x_{2k}) \geq 1.5$, $\sigma_{vk} = \tilde{\sigma}_v \cdot \{15 - F\left(g(x_{1k}, x_{2k})\right)\} / \{15 - F(1.5)\}$. Otherwise, let $\sigma_{vk} = \tilde{\sigma}_v$.

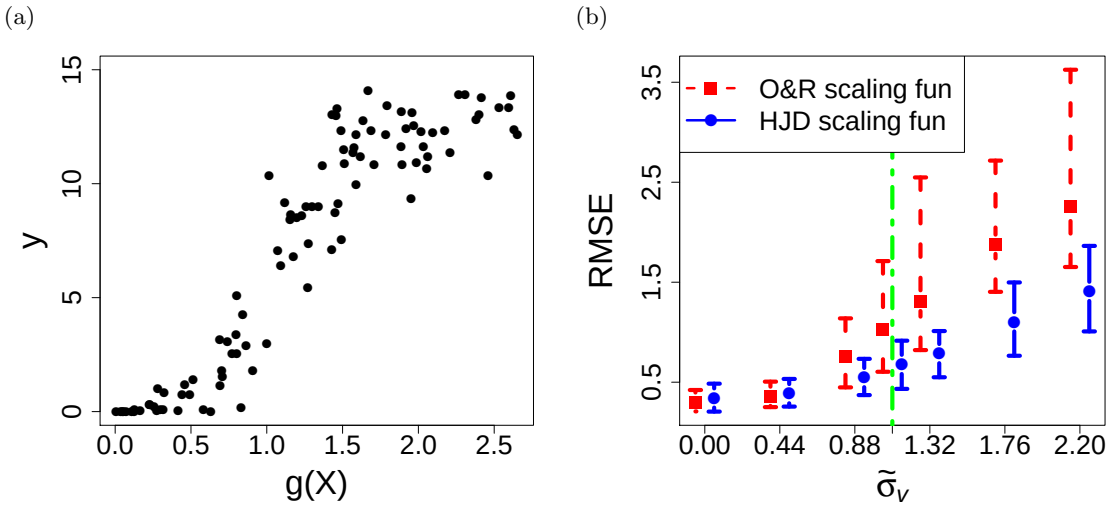
3: Obtain the output values according to $\tilde{y}_k = \phi(x_{1k}, x_{2k}) - u_k + v_k$. To ensure the observed outputs are in the range of [0, 15] which is the bounded range of $\phi$, let $y_k = |\tilde{y}_k|/5$ if $\tilde{y}_k < 0$, $y_k = 15 - |\tilde{y}_k - 15|/5$ if $\tilde{y}_k > 15$, and $y_k = \tilde{y}_k$ otherwise.

With the simulated data, we can adjust $\tilde{\sigma_u}$ and $\tilde{\sigma_v}$ values used so that the simulated mean and standard deviation of the composite error match, respectively, those of the real wind data. This

allows us to have an idea about how much the two types of randomness are there in the system. The mean and standard deviation of the composite error in the real data are 2.09 and 1.61, respectively, after rescaling for similarity to the O&R's DGP.

Figure 8(a) shows the scatter plot of a set of simulated data. If we choose $\tilde{\sigma}_u = 2.2$ and $\tilde{\sigma}_v = 1.1$ in our simulation, we produce a data set of which the averaged mean inefficiency and the standard deviation of composite errors are similar to those of the wind data, respectively. We then keep $\tilde{\sigma}_u = 2.2$, while varying $\tilde{\sigma}_v$ from zero to 2.2 (so that the signal-to-noise ratio is 1). The comparison results, when applying both HJD and O&R, are shown in Figure 8(b). The dotted line corresponds to $\tilde{\sigma}_v = 1.1$, indicating where the characteristics of our wind turbine data would fall in this analysis. Figure 8 shows that HJD is much more effective for the frontier estimation considering the noise level observed in the wind data. In fact, the data used in the previous example were measured during the first year of operation of that particular wind turbine. As a wind turbine operates for a longer period of time, the noise level in data tends to increase (see Figure 9b in Staffell and Green 2014), suggesting that HJD is better suited for estimating future power curves also.

**Figure 8** Simulation results: (a) scatter plot of simulated data mimicking wind turbine data and (b) comparison results using the simulated data with various scales of noise.




*Note.* For (b), $\tilde{\sigma}_u = 2.2$ with varying $\tilde{\sigma}_v$ indicated along the $x$-axis.

## 8. Conclusions

This paper presents a new RUP law-satisfying estimator, which also allows heteroscedasticity in both the random noise and inefficiency modeling. This new estimator is a major extension of the existing deterministic RUP law-satisfying estimators that assume data as noise free. Our studies using simulated data substantiate our claim that as soon as observations become noisy, the new stochastic estimator outperforms its deterministic counterpart. The advantage of the stochastic estimator becomes clear when the scale of noise is large.

We apply the resulting estimator to wind turbine data and estimate the average-practice power curve as well as its frontier function. Fitting these production functions sheds lights on the relative scale of the systematic inefficiency of a wind turbine and the random noise in its power production data. It turns out that the noise scale in the turbine data is relatively high; in the particular example we study, the signal-to-noise ratio is about two, a circumstance apparently demanding the use of the stochastic estimator developed in this paper. Bear in mind that our turbine data came from a turbine during its first year's operation. For turbines further into their service life, it is anticipated that the noise level will only go higher, resulting in even poorer performance of the deterministic estimators.

Finally, we feel that future research is needed to relax the input homotheticity assumption. The use of this assumption significantly simplifies the structure of the production function. Making no assumptions about the structure of inputs is not desirable because the estimation procedures become extremely flexible. To make the resulting methodology broadly applicable, a potentially better compromise between flexibility and efficiency may be the adaptation of the ray-homothetic structure described in Fare and Shephard (1977). Developing the ray-homothetic structure into a stochastic estimator represents a considerable improvement.

## Acknowledgments

Hwangbo and Ding's research was partially supported by the National Science Foundation under grant no. CMMI-1300560.

## References

Afriat SN (1967) The construction of a utility function from expenditure data. *International Economic Review* 8(1):67–77.

Aguilera N, Forzani L, Morin P (2011) On uniform consistent estimators for convex regression. *Journal of Nonparametric Statistics* 23(4):897–908.

Banker RD, Maindiratta A (1992) Maximum likelihood estimation of monotone and concave production frontiers. *The Journal of Productivity Analysis* 3(4):401–415.

Bogetoft P, Tama JM, Tind J (2000) Convex input and output projections of nonconvex production possibility sets. *Management Science* 46(6):858–869.

Chambers RG (1988) *Applied Production Analysis: A Dual Approach* (Cambridge University Press, Cambridge).

Chambers RG, Mitchell T (2001) Homotheticity and non-radial changes. *Journal of Productivity Analysis* 15(1):31–39.

Fare R, Shephard RW (1977) Ray-homothetic production functions. *Econometrica* 45(1):133–146.

Førsund FR, Hjalmarsson L (2004) Are all scales optimal in DEA? Theory and empirical evidence. *Journal of Productivity Analysis* 21(1):25–48.

Frisch R (1965) *Theory of Production* (D. Reidel Publishing Company, Dordrecht-Holland).

Hall P, Simar L (2002) Estimating a change point, boundary, or frontier in the presence of observation error. *Journal of the American Statistical Association* 97(458):523–534.

Hanson DL, Pledger G (1976) Consistency in concave regression. *The Annals of Statistics* 4(6):1038–1050.

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

Holloway CA (1979) On the estimation of convex functions. *Operations Research* 27(2):401–407.

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

Lee C, Johnson AL, Moreno-Centero E, Kuosmanen T (2013) A more efficient algorithm for Convex Nonparametric Least Squares. *European Journal of Operational Research* 227(2):391–400.

Lim E, Glynn PW (2012) Consistency of multidimensional convex regression. *Operations Research* 60(1):196–208.

Mammen E (1991) Nonparametric regression under qualitative smoothness assumptions. *The Annals of Statistics* 19(2):741–759.

Mammen E, Thomas-Agnan C (1999) Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics* 26(2):239–252.

Olesen OB, Petersen NC (2013) Imposing the regular ultra passum law in DEA models. *Omega* 41(1):16–27.

Olesen OB, Ruggiero J (2014) Maintaining the regular ultra passum law in data development analysis. *European Journal of Operational Research* 235(3):798–809.

Park BU, Simar L (1994) Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association* 89(427):929–936.

Parzen E (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3):1065–1076.

Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27(3):832–837.

Ruggiero J (1996) On the measurement of technical efficiency in the public sector. *European Journal of Operational Research* 90(3):553–565.

Seijo E, Sen B (2011) Nonparametric least squares estimation of a multivariate convex regression function. *The Annals of Statistics* 39(3):1633–1657.

Silverman BW (1986) *Density estimation for statistics and data analysis* (Chapman and Hall, New York).

Staffell L, Green R (2014) How does wind farm performance decline with age? *Renewable Energy* 66:775–786.

# Supplementary Material

## EC.1. Extension to the Multiple Outputs Setting

By assuming separability between inputs and outputs, multiple outputs can be represented by one dimensional aggregate output. As we aggregate multiple inputs into an aggregate input, similar approaches can be applied to obtain an aggregate output: 1) select subset of population based on the density estimate of inputs and 2) apply CNLS to estimate a level set of base input. Then, the subsequent procedures of inflection point estimation and production frontier function estimation can be applied based on this two dimensional aggregate input-aggregate output space.

## EC.2. Proofs

### EC.2.1. Proof of Lemma 1

*Proof of Lemma 1* Suppose that $x^* \in [\min_k x_k, \eta_{c^--1}]$. According to the RUP law, $\psi$ is (strictly) concave in $[\eta_{c^--1}, \max_k x_k]$ (excluding $\eta_{c^--1}$). By the concavity of $\psi$ and the conditions in (7),

$$\psi'(\eta_{c^-}) \leq \frac{\xi_{c^-} - \xi_{c^--1}}{\eta_{c^-} - \eta_{c^--1}} < \frac{\xi_{c^+} - \xi_{c^-}}{\eta_{c^+} - \eta_{c^-}} \tag{EC.1}$$

Furthermore, because $\psi$ is concave in $[r_{c^-}, r_{c^+} + h)$, as $n_b \to \infty$ for $b = c^-, c^+$, $\xi_{c^+}$ and $\xi_{c^-}$ underestimate $\psi(\eta_{c^+})$ and $\psi(\eta_{c^-})$, respectively, i.e. $\xi_{c^+} = \psi(\eta_{c^+}) - \delta_{c^+}$ and $\xi_{c^-} = \psi(\eta_{c^-}) - \delta_{c^-}$ where $\delta_{c^+}, \delta_{c^+} > 0$ are the errors of the estimator $\xi_{c^+}$ and $\xi_{c^-}$, respectively. Note that stronger curvature of $\psi$ in $[r_{c^+}, r_{c^+} + h)$ than in $[r_{c^-}, r_{c^-} + h)$ due to the decreasing second derivative of $\psi$ leads to $\delta_{c^+} > \delta_{c^-}$. Then, (EC.1) becomes

$$\psi'(\eta_{c^-}) < \frac{\psi(\eta_{c^+}) - \psi(\eta_{c^-})}{\eta_{c^+} - \eta_{c^-}} - \frac{\delta_{c^+} - \delta_{c^-}}{\eta_{c^+} - \eta_{c^-}} < \frac{\psi(\eta_{c^+}) - \psi(\eta_{c^-})}{\eta_{c^+} - \eta_{c^-}}$$

indicating that $\psi$ is convex at least in the interval of $[\eta_{c^-}, \eta_{c^+}] \subset (\eta_{c^--1}, \max_k x_k]$. This contradicts the assumption that $\psi$ is strictly concave in $(\eta_{c^--1}, \max_k x_k]$. Therefore, $x^* \notin [\min_k x_k, \eta_{c^--1}]$. Similarly, we can show that $x^* \notin [\eta_{c^++1}, \max_k x_k]$. These together prove the desired statement. □

### EC.2.2. Proof of Theorem 1

*Proof of Theorem 1* Since $x_k$'s are uniformly distributed in $[\inf_l r^l_{c^- -1}, \sup_l r^l_{c^+ +1} + h)$ and $n_b \to \infty$, we have $\eta^l_b \to r^l_b + h/2$ for $b = c^-(l) - 1$, $c^-(l)$, $c^+(l)$, $c^+(l)+1$ and for $\forall\, l$. Note that by construction, $|r^l_{c^- -1} - r^{l'}_{c^- -1}| \le h$, so $|\eta^l_{c^- -1} - \eta^{l'}_{c^- -1}| \le h$ for $\forall\, l$ and $\forall\, l'$. This also implies that $\eta^l_{c^-} \ge \eta^{l'}_{c^- -1}$ for $\forall\, l, l'$.

Consider a sequence $\left(\sup\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\} - \inf\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\}\right)_{n_{rep}}$, which is increasing and bounded above by $h$. Thus, as $n_{rep} \to \infty$, the sequence converges to $h$. Thanks to the fact that $\eta^l_{c^-} - \eta^l_{c^- -1} \to r^l_{c^-} + h/2 - \left(r^l_{c^- -1} + h/2\right) = \left(r^l_{c^- -1} + h\right) - r^l_{c^- -1} = h$,

$$\left(\inf\left\{\eta^l_{c^-} :\ l = 1, \ldots, n_{rep}\right\} - \sup\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\}\right)_{n_{rep}}$$
$$\to \left(h + \inf\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\} - \sup\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\}\right)_{n_{rep}} \to 0.$$

Similarly,

$$\left(\inf\left\{\eta^l_{c^+ +1} :\ l = 1, \ldots, n_{rep}\right\} - \sup\left\{\eta^l_{c^+} :\ l = 1, \ldots, n_{rep}\right\}\right)_{n_{rep}} \to 0.$$

Then, by Lemma 1, we have

$$x^* \in \bigcap_{l=1}^{n_{rep}} \left(\eta^l_{c^- -1}, \eta^l_{c^+ +1}\right) = \left(\sup\left\{\eta^l_{c^- -1} :\ l = 1, \ldots, n_{rep}\right\}, \inf\left\{\eta^l_{c^+ +1} :\ l = 1, \ldots, n_{rep}\right\}\right)$$
$$= \left(\inf\left\{\eta^l_{c^-} :\ l = 1, \ldots, n_{rep}\right\}, \sup\left\{\eta^l_{c^+} :\ l = 1, \ldots, n_{rep}\right\}\right) = \left(\inf \mathcal{C}^-_x, \sup \mathcal{C}^+_x\right). \quad \square$$

## EC.3. The Argument to Demonstrate (10h) and (10j) are Sufficient for Imposing the Weak Essentiality Axiom

We define production function estimate $\hat{\psi}$. Since the fitted output $\hat{y}_{k'}$'s are point estimates, any continuous function passing $(x_{k'}, \hat{y}_{k'})$ pairs can represent $\hat{\psi}$ with an equivalent objective function value to (10). To address the non-uniqueness issue, Kuosmanen (2008) has constructed an explicit representor function, so for the concave region,

$$\tilde{\psi}(x) = \min_{k' \in \{k^*, \ldots, n+1\}} \left\{\hat{\beta}_{k',0} + \hat{\beta}_{k',1} x\right\}.$$

Kuosmanen has also established the tightest possible lower bound and upper bound of the explicit representor function $\tilde{\psi}(x)$. Kuosmanen and Kortelainen (2012) show that the tightest possible lower bound minimizes the sample variance of deviations and suggest using the tightest possible lower bound as CNLS estimate:

$$\tilde{\psi}_{\min}(x) = \min_{\beta_0 \in \Re,\, \beta_1 \in \Re_+} \{\beta_0 + \beta_1 x : \ \beta_0 + \beta_1 x_{k'} \geq \hat{y}_{k'},\ k' = k^*, \ldots, n+1\}. \tag{EC.2}$$

In (EC.2), the parameters $\beta_0$ and $\beta_1$ are reestimated, so they can be distinct from $\hat{\beta}_{k',0}$ and $\hat{\beta}_{k',1}$ obtained from (10). However, note that $\tilde{\psi}_{\min}(x_{k'}) = \tilde{\psi}(x_{k'}) = \hat{y}_{k'}$ for the observed $x_{k'}$'s. The discrepancy between $\tilde{\psi}(x)$ and $\tilde{\psi}_{\min}(x)$ only occurs between two successive observations $x_{k'}$ and $x_{k'+1}$ where $\hat{\beta}_{k',0} \neq \hat{\beta}_{k'+1,0}$ and $\hat{\beta}_{k',1} \neq \hat{\beta}_{k'+1,1}$. In this case, $\tilde{\psi}(x)$ extends the two hyperplanes, $(\hat{\beta}_{k',0}, \hat{\beta}_{k',1})$ and $(\hat{\beta}_{k'+1,0}, \hat{\beta}_{k'+1,1})$, up to the point where they meet while $\tilde{\psi}_{\min}(x)$ imposes another hyperplane passing through both of $(x_{k'}, \hat{y}_{k'})$ and $(x_{k'+1}, \hat{y}_{k'+1})$. So, $\tilde{\psi}_{\min}(x)$ can be interpreted as a function connecting all $(x_{k'}, \hat{y}_{k'})$'s and extending at the boundaries of support of the function.

For the convex region, we can reverse the direction of inequalities in (EC.2) and take the maximum instead of the minimum. Then, our estimator $\hat{\psi}(x)$ for the S-shaped production function is

$$\hat{\psi}(x) = \begin{cases} \max\limits_{\beta_0 \in \Re_-,\, \beta_1 \in \Re_+} \{\beta_0 + \beta_1 x : \ \beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'},\ k' = 1, \ldots, k^*\}, & \text{if } x \leq x_{k^*} \\ \min\limits_{\beta_0 \in \Re,\, \beta_1 \in \Re_+} \{\beta_0 + \beta_1 x : \ \beta_0 + \beta_1 x_{k'} \geq \hat{y}_{k'},\ k' = k^*, \ldots, n+1\}, & \text{if } x \geq x_{k^*}. \end{cases} \tag{EC.3}$$

With the production function estimate $\hat{\psi}$ defined on (EC.3), we show sufficiency of (10h) and (10j) for the imposition of the origin.

Proposition EC.1. *Production function estimate $\hat{\psi}$ passes through the origin if $\hat{\beta}_{k',0} \leq 0$ for $k' = 1, \ldots, k^*$ and $\hat{y}_{k'} \geq 0$ for $\forall k' = 1, \ldots, n+1$, in addition to other constraints in* (10).

*Proof of Proposition EC.1* (i) Assume that $\hat{y}_1 = 0$, i.e. the minimum of point-wise functional estimate is zero. If $x_1 = 0$, the statement always holds. Thus, consider the case of $x_1 > 0$. Note that $0 < x_1 < x_{k^*}$, otherwise $x_{k^*}$ cannot be an inflection point estimate. Then,

$$\hat{\psi}(x_1) = \max_{\beta_0 \in \Re_-,\, \beta_1 \in \Re_+} \{\beta_0 + \beta_1 x_1 : \ \beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'},\ k' = 1, \ldots, k^*\} = \hat{y}_1 = 0$$

which implies there exist $\beta_0 = \beta_1 = 0$ such that $\beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'}$ for $k' = 1, \ldots, k^*$. So,

$$\hat{\psi}(0) = \max_{\beta_0 \in \Re_-, \beta_1 \in \Re_+} \{\beta_0 : \ \beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'}, \ k' = 1, \ldots, k^*\} = 0$$

because $\beta_0$ is bounded above by 0.

(ii) Suppose that $\hat{y}_1 > 0$. Because $\hat{\beta}_{1,0} \leq 0$, $x_1 > 0$ and $\hat{\beta}_{1,1} > 0$ whenever $\hat{y}_1 > 0$. If $\hat{\beta}_{1,0} = 0$, there exist $\beta_0 = 0$ and $\beta_1 = \hat{\beta}_{1,1} > 0$ such that $\beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'}$ for $k' = 1, \ldots, k^*$. Thus, $\hat{\psi}(0) = 0$.

Now, the only case left is when $\hat{\beta}_{1,0} < 0$. Let $\tilde{\beta}_{0,1} = (\hat{y}_1 - 0)/(x_1 - 0)$. Then, $\tilde{\beta}_{0,1} x_1 = \hat{y}_1 = \hat{\beta}_{1,0} + \hat{\beta}_{1,1} x_1$. Since $\hat{\beta}_{1,0} < 0$, $\tilde{\beta}_{0,1} < \hat{\beta}_{1,1}$. Connecting $(0,0)$ and all $(x_{k'}, \hat{y}_{k'})$'s, thus, form a piece-wise linear and convex curve. We can write

$$\hat{y}_{k'} \geq \hat{y}_1 + \tilde{\beta}_{0,1}(x_{k'} - x_1) = \hat{y}_1 - \tilde{\beta}_{0,1} x_1 + \tilde{\beta}_{0,1} x_{k'} = 0 + \tilde{\beta}_{0,1} x_{k'}, \ \text{for } k' = 1, \ldots, k^*.$$

Therefore, $\beta_0 = 0$ and $\beta_1 = \tilde{\beta}_{0,1}$ satisfies $\beta_0 + \beta_1 x_{k'} \leq \hat{y}_{k'}$ for $k' = 1, \ldots, k^*$, so $\hat{\psi}(0) = 0$. $\square$

## References

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

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.