PERFORMANCE EVALUATION OF WIND POWER SYSTEMS BASED ON PRODUCTION ECONOMICS THEORY

A Dissertation

by

HOON HWANGBO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yu Ding
Committee Members,	Andrew L. Johnson
	Lewis Ntaimo
	Chanan Singh
Head of Department,	Mark Lawley

August 2017

Major Subject: Industrial Engineering

ABSTRACT

The research in this dissertation addresses the issues in the performance evaluation of wind power systems under commercially operating circumstances. Such an evaluation is critical to a wide range of decisions including operations and maintenance planning, reliability assessment, asset procurement, and system designs. However, accurate evaluation is excessively challenging due to the unknown causal relationship between wind input and power output, the dependency of power output on numerous uncontrollable factors, and the high level of uncertainty observed in power output. While addressing these challenges, we develop a new performance measure based on production economics theories and propose effective methodologies for evaluating the performance of wind power systems. By doing so, this dissertation study aims to improve the practice of performance evaluation in the wind industry.

We define an efficiency metric analogous to productive efficiency, which requires estimating a performance benchmark, i.e., the performance referring to 100% efficiency. For the performance benchmark, we develop a stochastic nonparametric estimator maintaining S-shape, the typical shape observed in the wind input-power output relationship. When applying the efficiency metric for comparing performance under different scenarios, other environmental factors need to be controlled for, as their difference could produce a difference in power output. We devise a covariate density matching method that selects subsets of data for which probability densities of the environmental factors are comparable; evaluating only these subsets, then, ensures a fair comparison. We further investigate wake situations in which the operation of a turbine could cause a significant power deficit on its neighboring turbines. In the presence of the performance benchmark introduced earlier,

we can model the power deficit as a non-negative term subtracted from the benchmark. Based on this model setup, we develop a spline model with a non-negativity constraint imposed for characterizing such a wake effect.

When each of the proposed methods is applied to operational wind data, the respective results demonstrate that each of the methods outperforms its competitive alternatives in terms of estimation and/or prediction accuracy. This suggests that the methods can reduce the unaccounted uncertainty in power output and thus provide better insight into the performance of wind power systems.

DEDICATION

This dissertation is dedicated to my family who have supported me in various ways until the completion of the work.

First of all, I dedicate this dissertation to my wife, Bohyun Jung, for sharing all my joys and sorrows at the closest distance and supporting me in any and every possible way with full of love and patience.

I also dedicate this dissertation to my son, Daniel Young Hwangbo, who has made my life full of excitement.

Finally, I dedicate this work to my parents, my mother-in-law, and my sister for their constant prayer and wholehearted support.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my advisor, Prof. Yu Ding, for his continuous and incomparable support of all my development as a professional during my Ph.D. study. He has provided me invaluable opportunities for cultivating my skill sets associated with research, teaching, student mentoring, and collaboration, while constantly motivating me with challenging and constructive comments. While bearing with me whenever I encountered a problem, he has been always willing to help and has shown his confidence in me. This mentorship has enabled the completion of this dissertation, and I cannot imagine such completion without Dr. Ding's dedicating help. I believe working with him was the best decision I have ever made for my Ph.D. study.

I would also like to thank Prof. Andrew L. Johnson, one of my committee members, for educating me and helping me accumulate the knowledge of production economics that established the backbone of this dissertation. He provided me useful sources of knowledge, suggested promising directions to proceed, and helped improve many of the technical details. I also owe my sincere gratitude to the other committee members, Prof. Lewis Ntaimo and Prof. Chanan Singh, for their insightful comments and constant encouragement, which not only broadened my perspective improving this dissertation study but also inspired the direction for some future study.

Furthermore, I thank my labmates, Arash Pourhabib, Yanjun Qian, Ahmed Aziz Ezzat, Erika Sy, Imtiaz Ahmed, David Pérez, Qian Wu, and Doowon Choi, for providing me a stimulating environment, constructive discussion, and daily enjoyment that have expedited the completion of my Ph.D. study. I believe that all the experience with the labmates is a valuable asset and will help me to succeed in my future career, especially in building my

professional network.

My thanks are also due to two undergraduates who were under my supervision, Briana Niu and Victor Gálvez Yanjari, for their enthusiasm for working with me and solving a given research problem that ultimately resulted in submission of the research work to a well-known journal. The research work has provided me useful insight into the significance of this dissertation study. My sincere thanks also goes to Prof. Li Zeng for supporting the discussions associated with the undergraduate research work and providing productive comments.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Yu Ding, Professor Andrew L. Johnson and Professor Lewis Ntaimo of the Department of Industrial & Systems Engineering and Professor Chanan Singh of the Department of Electrical & Computer Engineering.

The analyses depicted in Chapter IV were conducted in part by Mingdi You, Professor Eunshin Byon, Professor Jionghua (Judy) Jin of the Department of Industrial & Operations Engineering, University of Michigan, Ann Arbor, and Professor Giwhyun Lee in the Korea Army Academy at Yeong-Cheon, South Korea.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was partially supported by the U.S. National Science Foundation under grant no. CMMI-1300560.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION*

I.1 Motivation

A well-known concern for future energy shortages and adverse environmental impacts of fossil-based energy sources has driven an emphasis on renewable energy and has motivated extensive research on the implementation and improvement of renewable energy technology. Among various renewable energy sources, wind energy is one of the most promising. In 2015, wind energy supplied 4.7% of the total electricity generated in the U.S. (AWEA, 2016). More significantly, the electricity generated from wind energy has increased almost tenfold in the past decade, from 18 million megawatt-hours (MWh) in 2005 to 191 million MWh in 2015 (Energy Information Administration, 2016). In addition, the Department of Energy envisions that wind energy could supply 10% of the nation's electricity by 2020, 20% by 2030, and 35% by 2050 (DOE, 2015).

With the rapid promotion of wind energy and the significant increase in its capacity, evaluating wind turbine performance in terms of its power production under commercially operating circumstances becomes increasingly important. A worthwhile evaluation identifies the best practice of both design and operation in the wind industry, helping decide when to conduct a major maintenance task, which turbine brand runs more efficiently, and whether a retrofitting upgrade is worth the trouble to have it installed.

The evaluation, though, is not an easy task. A simple comparison of power output often misleads about the underlying performance of wind turbines as the power output strongly depends on wind input that is highly variable and uncontrollable in its nature.

*Part of this chapter is reprinted with permission from Hwangbo, H., Johnson, A. L., and Ding Y. (2017). A production economics analysis for quantifying the efficiency of wind turbines. *Wind Energy* (in press) http://dx.doi.org/10.1002/we.2105, Copyright © 2017 by John Wiley & Sons, Ltd.

For example, a wind turbine may generate a considerable amount of power output for the period of evaluation, but this strong performance could be due to the strong wind existed in that evaluation period rather than anything unique about its design or operation. As wind input changes all the time, performance evaluation will have to account for the difference in wind input as well.

In the next sections, we will first explain some of the background regarding wind power production, particularly the current practice of turbine performance evaluation. Then, we will proceed to elaborate why production economics theory is relevant and can address the technical challenge in wind turbine performance evaluation. Finally, we outline the research carried out in this dissertation study and the structure of this dissertation.

I.2 Background of wind power production and current efficiency metrics

The amount of power a wind turbine can extract is defined by the power production equation (Boukhezzar et al., 2006), which has the form of

$$y = \frac{1}{2}\rho A V^3 C_p, \tag{I.1}$$

where y is the extractable wind power, V denotes wind speed, ρ denotes air density, and $A = \pi R^2$ is the rotor swept area for a rotor of radius R. The C_p, known as the power coefficient, is not a constant but a function. Its functional form is, however, unknown and generally includes pitch angle and tip speed ratio that depends on wind speed as its arguments. While A remains constant for a specific turbine model, the other terms of ρ, V, and C_p are variable.

For performance evaluation of wind turbines, the International Electrotechnical Commission (IEC) (IEC12.1, 2005) recommends using (i) annual energy production (AEP), (ii) power curve, or (iii) power coefficient curve. The AEP has a crucial drawback if used for the performance evaluation since it only quantifies an amount of power output and does

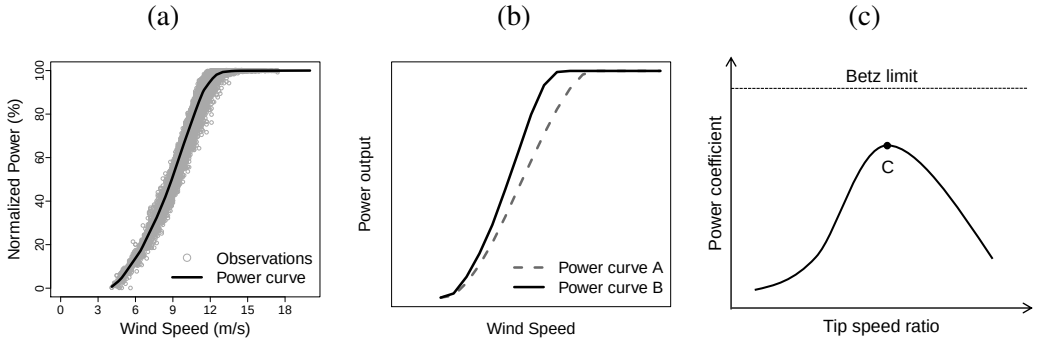

Figure I.1: Illustration of turbine performance evaluation: (a) A scatter plot of wind speed and power data, and the estimated power curve; (b) Two power curves indicating relative efficiencies of wind turbines in which curve B suggests a higher productive efficiency; (c) Power coefficient curve and the Betz limit.

not account for the difference in wind input. On the other hand, power curves and power coefficient curves are capable of describing power output as a function of wind input and hence more appropriate for the performance evaluation.

Power curve is a functional curve, showing the relationship between wind speed input and wind power output. As shown in Figure I.1(a), power curves are typically estimated from the measured power output data and wind speed data by finding a smooth curve minimizing the squared residual errors. Then, the resulting curve passes through the middle of the data, representing the average power production of a wind turbine. Once the power curves are estimated, the relative positions on the power curve plot may suggest relative productive efficiency (see Figure I.1(b)).

In the same way as the estimation of a power curve, the estimation of a power coefficient curve (often plotted against the tip speed ratio) averages the power coefficients computed from the observational data through the power production equation. In practice, the largest value on the curve (for instance, point C in Figure I.1(c)), as the representative of the whole curve, is used for quantification of the aerodynamic efficiency (Homola et al.,

2012; Eriksson et al., 2008; Krogstad and Lund, 2012). Hereafter, we will refer to the representative point as a power coefficient unless otherwise stated.

The performance evaluation mechanisms are often translated to the following numerical efficiency metrics: (i) capacity factor, (ii) power coefficient, or (iii) power generation ratio. An efficiency metric is usually calculated as a ratio of a performance measure over a benchmark quantity. This is true for a capacity factor and a power generation ratio. However, in practice, the power coefficient is used as an efficiency metric as is without being normalized by its theoretical benchmark.

The capacity factor is the ratio of the observed power output over a turbine's maximum power capacity (Wikipedia, 2017). This calculation assumes a turbine is operating at its full capacity all the time. As such, the capacity factor significantly underestimates the ground truth since an achievable power output could be much lower than the maximum capacity depending on the wind speed (Figure I.1(a)).

In the case of the power coefficient, practitioners use the theoretical upper limit, known as the Betz limit (=0.593; Betz 1966), as the performance benchmark. Yet, this upper limit is not practically achievable; in general, the power coefficients estimated are below 0.45. So when normalizing by the Betz limit, the corresponding efficiency measure never approaches one, which could be the reason why a normalization is not used. A more crucial limitation for the power coefficient is that the efficiency quantification is based on a point representation of the power coefficient curve, so that it cannot differentiate two turbines that have the same peak power coefficient values but different power coefficient curves.

The power generation ratio uses, as the performance benchmark, the expected power output obtained from a nominal power curve. A nominal power curve provided by turbine manufacturers for performance guarantee is an averaged representation of the power performance as it is a power curve. As such, an observed power output could exceed

Capacity Factor	Power Coefficient	Power Generation Ratio
$\frac{Observed\ power\ output}{Maximum\ power\ capacity}$	$\frac{Observed\ power\ output}{Energy\ available\ in\ the\ wind}$	$\frac{Observed\ power\ output}{Expected\ nominal\ power\ output}$

Figure I.2: Three efficiency metrics commonly used in the wind industry.

the averaged representation, so the power generation ratio frequently goes above one (or 100%). Furthermore, a nominal power curve does not reveal site-specific conditions and is usually different from operational power curves (Gill et al., 2012; Papatheou et al., 2015; Uluyol et al., 2011; Khalid and Savkin, 2012; Hayes et al., 2011), but such discrepancy is not taken into account in the calculation of the power generation ratio.

Figure I.2 summarizes how to calculate the three metrics and how they differ from each other. An important observation from the figure is that all the metrics share the same numerator but use a different denominator, implying that there is no consensus about the benchmark quantity to be used. In fact, such performance benchmark is difficult to characterize. The need for proper performance benchmark motivates us to look into the field of production economics where efficiency analysis and thereby estimation of performance benchmark is one of the main interests. We describe some production economics theory and its relevance in more detail in the following section.

I.3 Production economics theory and its relevance

Production economics analysis concerns the efficiency of individual production systems that utilize some inputs to produce a certain type of output, in terms of their input usage and/or production output relative to some benchmark available for those sharing a common technology, e.g., production systems in the same industry. So-called production function describes output-oriented benchmark (as a function of inputs) and hence plays a critical role in the efficiency analysis of production systems. For better understanding of

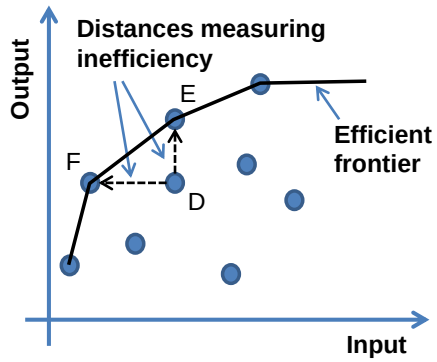

Figure I.3: Production data and efficient frontier.

its basic idea, we proceed with the following simple example.

Consider a set of production units (e.g., a wind farm) using x input (e.g., investment in a wind energy project) and producing y output (e.g. revenue from power generation). We can create a scatter plot of many x-y data pairs coming from different production units or the same production unit but over different periods (see Figure I.3). Assuming no measurement errors associated with x and y, a common estimator in production economics, Data Envelopment Analysis (DEA; Banker et al. 1984), estimates an efficient frontier by enveloping all the observations.

The concept of an efficient frontier is understood as follows: a production unit whose input-output is on the frontier is more efficient than the production units whose input-output is being enveloped by the frontier. Consider observation D. Using the same input, the production unit associated with D produces less output than the production unit associated with point E; while to produce the same output, the production unit associated with D needs more input than the production unit associated with point F. So the production unit associated with D must be inefficient.

The efficient frontier is also called production function, denoted by $f(x)$. The production function characterizes producible output given input x in the absence of inefficiency.

Using the production function, the output of the inefficient production unit D can be expressed as

$$y_D = f(x_D) - u_D, \tag{I.2}$$

where $u_D \geq 0$ denotes the systematic inefficiency.

A wind turbine utilizes wind input and produces power output. As such, we may construct a production function of a wind turbine, i.e., by setting x = wind input and y = power output in the above example. In fact, a similar relationship has been already established in the wind energy literature in the form of power curve although it is an average representation. Our objective is then to estimate a production function of a wind turbine or, equivalently, the best-practice power curve that can serve as the performance benchmark.

To estimate a production function, certain assumptions are made restricting the shape of the frontier. The most common assumption is that the frontier forms a monotone increasing concave function consistent with basic stylized characteristics of production (Varian, 1982). When the data are assumed noise free, the tightest boundary enveloping all observations and maintaining the monotonicity and concavity is a piece-wise linear function as shown in Figure I.3.

Convex or concave piecewise linear methods assuming noise-free data encounter some problems when applied to wind turbine data. The first is that the wind turbine data, like all other physical measurements, are inevitably contaminated by noises. The second difference is that the shape of the wind-power scatter plot is not concave. Instead, the data appears to follow an S-shape, as shown in Figure I.4, comprising a convex region, followed by a concave region, and the two segments of the curve are connected at an inflection point.

In production economics, the need to model noise is well established, promoting the subfield of stochastic frontier analysis (SFA; Aigner et al. 1977). The SFA model includes

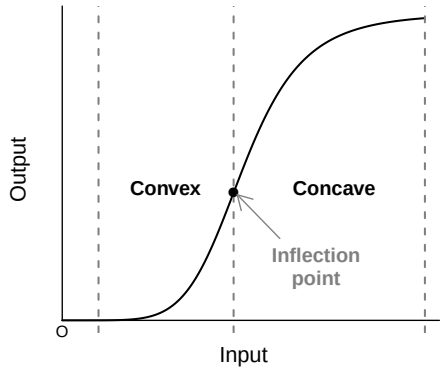

Figure I.4: A 2-dimensional S-shaped curve: a region where the production function is convex followed by a region where the function is concave, connected with an inflection point.

a random noise term v to Equation (I.2), as follows.

$$y = f(x) - u + v. \tag{I.3}$$

While the current SFA research considers the noise effect in observational data, researchers typically do not address the second difference mentioned above, namely, the S-shape exhibited in the wind turbine data. Instead, researchers typically rely on parametric functional forms, such as Cobb-Douglas, that need not satisfy the S-shape for any parameter values. In fact, the S-shape constraint corresponds to the Regular Ultra Passum (RUP) law (Hackman, 2008) in economics that is motivated by production units having an increasing marginal rate of productivity followed by a decreasing rate of marginal productivity. As shown in Figure I.1, it is obvious that wind turbine power curves also satisfy the RUP law.

Very few production function estimators impose the RUP law explicitly. The exception is the DEA-type estimator developed by Olesen and Ruggiero (2014). Unfortunately, the frontier analysis employed by Olesen and Ruggiero (2014) is still deterministic, enveloping all observations from above, and consequently suffers from the overestimation that all

other deterministic production function estimators suffered.

In this dissertation, we first use shape constrained functional estimation to model the typical shape of noisy wind data, which is consistent with the RUP law. Given the performance benchmark represented by such a production function, we formally define a new efficiency metric for a wind turbine system. The newly defined efficiency metric is used to study a wind turbine's performance change over several years. The development of the performance benchmark also facilitates the modeling of wake effect in a single wake setting (i.e., a pair of two turbines) and makes the subsequent wake effect estimation easier to be carried out.

I.4 Organization of this dissertation

This dissertation is organized as follows. Chapter II develops a production function estimator satisfying the RUP law. To derive the estimator, we follow the sequential steps devised for the deterministic counterpart (Olesen and Ruggiero, 2014) but replace each of the steps by including a model of noise. Results from Monte Carlo simulations suggest that the proposed estimator outperforms the deterministic estimator as the noise level increases and also is more suitable for a typical wind data set considering the noise level revealed in the data.

Chapter III defines the efficiency of a wind power system and proposes a covariate density matching method that neutralizes the effect of environmental variables when comparing between a number of calculated efficiency. When applied to multiple operational turbine data, the resulting trajectories of the efficiency over several years all exhibit a similar trend to that obtained by an independent study, to some extent validating effectiveness of each other in performance monitoring.

Chapter IV presents how to model and estimate wake effect based on the existence of the performance benchmark represented by a production function. We also describe

important wake effect characteristics dictating the design of wind turbines and farms and their operations. For model estimation, we construct a spline-based model and fit the wake effect by using thin plate regression splines with non-negativity (TPRS-N), a novel approach proposed in this chapter. The results from its application to a set of operational turbine data suggest that the unique model structure allows the model to capture wake characteristics more effectively, providing wind energy practitioners with better insight.

Chapter V summarizes this dissertation and highlights its contributions. We also discuss potential extensions of this dissertation study.

CHAPTER II

POWER CURVE ESTIMATION: FUNCTIONAL ESTIMATION IMPOSING THE REGULAR ULTRA PASSUM LAW

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 imposing the RUP law between energy output and wind speed. The results demonstrate that the proposed estimator is well suited for engineering applications with a high degree of noise.

II.1 Introduction

The Regular Ultra Passum (RUP) law, first described by Frisch (1964), 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 wind turbine power curves 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), Kuosmanen (2001), 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 two-step procedure to generate an RUP-satisfying estimator that entails: 1) estimate a base isoquant; and 2) 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 is that 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 previously developed two-step procedure. This does not mean that our extension is straightforward; on the contrary, both steps need to be altered for an estimation procedure that includes a model of noise. Our estimator also 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. 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 chapter is organized as follows. Section II.2 introduces the production frontier model and the assumptions in our research development. Sections II.3 and II.4 present, respectively, the new version of base isoquant estimation and production function estimation, along with the adaptations needed to make the resulting estimator capable of modeling noise. Section II.5 discusses the Monte Carlo simulation results and demonstrates the estimator's improved performance over its deterministic counterpart when modeling noisy data. Section II.6 employs the resulting method to estimate a power curve for a wind turbine. Section II.7 concludes the chapter. All proofs and supplemental materials are available in Appendices.

II.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

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

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 μ, i.e., $E(u) = \mu$. We define production function ϕ 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., $g(c\boldsymbol{X}) = cg(\boldsymbol{X})$ for any positive scalar c. The production function ϕ is to be estimated by using n pairs of observations $\{(\boldsymbol{X}_1, y_1), \cdots, (\boldsymbol{X}_n, y_n)\}$.

We are interested in estimating a production function ϕ 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 qth 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 $\boldsymbol{X}_1$.

Functions that satisfy the RUP law are a subset of the quasiconcave functions defined by Fenchel and Blackett (1953), see also Diewert et al. (1981). However, this class of functions are particularly important because they satisfy standard axioms of production theory. Specifically, the regular ultra passum law assures that along every ray from the origin the production function has a monotonically increasing returns to scale region followed by a monotonically decreasing returns to scale region. The increasing returns to scale region is typically justified on the basis of returns to specialization, and the decreasing returns to scale region is motivated by decreasing marginal benefits of inputs due to for example increasing span of control or scarcity of highly productive activities.

Production functions defined over a single (aggregate) input that satisfy the RUP law

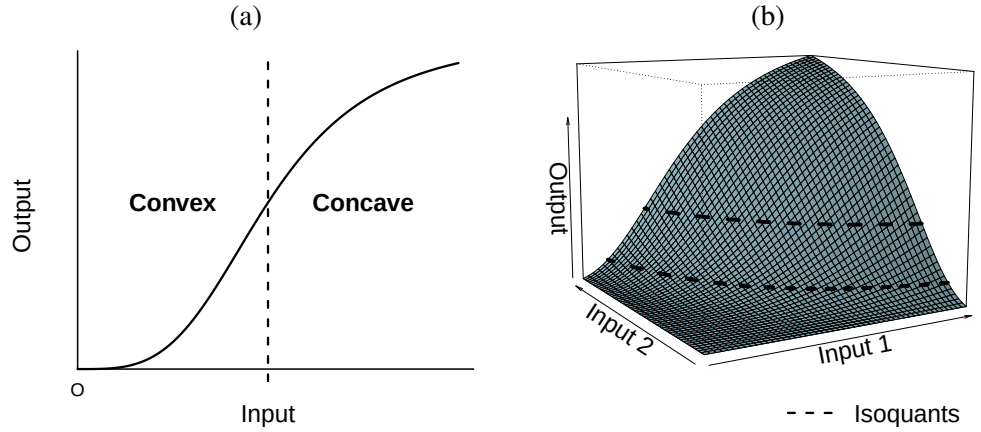

Figure II.1: Production functions satisfying the Regular Ultra Passum law: (a) S-shaped 2-dimensional production function that satisfies the Regular Ultra Passum law and has a region where the production function is convex followed by a region where the function is concave, (b) 3-dimensional production 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.

are often referred to as "S-shaped function" because of their shape. Figure II.1(a) shows an example of an S-shaped 2-dimensional production function, and Figure II.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 II.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 II.1(b), the space above the production function for low input levels is not a convex set (Frisch, 1964). To circumvent this difficulty, a common treatment,

as has been done by O&R, is to assume homotheticity on 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. We will follow the same strategy.

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 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; and Mammen (1991) who showed that the Hildreth-type estimator with a single input achieved the optimal nonparametric rate of convergence.

II.3 Input isoquant estimation and input aggregation

Input aggregation requires estimation of an input isoquant Isoq$L(y)$ defined as the boundary of a convex input set $L(y)$, namely

$$L(y) = \{\boldsymbol{X} \in \Re^p_+ : \boldsymbol{X} \text{ can produce } y\}, \tag{II.2}$$

$$\text{Isoq}L(y) = \{\boldsymbol{X} : \boldsymbol{X} \in L(y),\, \gamma\boldsymbol{X} \notin L(y),\, \gamma \in [0,1)\}. \tag{II.3}$$

Ideally, an input isoquant involves a set of input vectors that produces the same level of output, and it links the multi-dimensional inputs to the univariate output. Due to the monotonicity of a production function indicating that more input produces no less output, an isoquant associated with a higher output is placed further from the origin in the input space than the one associated with a lower output. Meanwhile, the homotheticity implies

that all isoquants share the same shape regardless of the associated output level and they are radially proportional to one another. One way to aggregate inputs for a homothetic function is to find a reference isoquant (base isoquant) revealing the common shape and to evaluate a relative usage of inputs to the one producing a base output y^*, output level associated with the reference isoquant. Specifically, a univariate score $\theta(\boldsymbol{X}, y^*)$ measures the relative usage by calculating the radial proportion of an input vector $\boldsymbol{X}$ to the reference isoquant evaluated at the same angular coordinate. The original multivariate problem with data pairs $(\boldsymbol{X}, y)$ then can be replaced by a two-dimensional problem considering (θ, y).

In the deterministic setting, Olesen and Ruggiero (2014) choose a base output of y_0, identify Isoq$L(y_0)$ by finding a deterministic boundary of the convex set $L(y_0)$, and calculate $\theta(\boldsymbol{X}, y_0)$ accordingly. In our problem, y is stochastic, differing from the production function ϕ by the inefficiency and random noise terms. This raises two questions: 1) what y value to use as the base output, and 2) how to estimate the base isoquant; both considering the presence of noise and inefficiency. In the following sections, we demonstrate how to aggregate multi-dimensional inputs by answering the two questions.

II.3.1 Base input isoquant estimation

For the discussion in this section, suppose that a base output y^* is given so the objective is to estimate the isoquant associated with y^*. Let $(\boldsymbol{X}_k, y_k)$ for $k = 1, \ldots, n$ denote the data pair we observe.

An isoquant defined in (II.3) can be rewritten based on the concept of a level set, for example

$$\text{Isoq}L(y^*) = \{\boldsymbol{X} : \phi(\boldsymbol{X}) = y^*\}, \tag{II.4}$$

where Isoq$L(y^*)$ is the isoquant associated with y^*. The equation (II.4) defines an isoquant as a set of input vectors that exactly produces a specific level of output in the absence of inefficiency and random noise. Without prior knowledge of the inefficiency and random

noise, isoquant estimation requires finding a reasonable subset of input vectors that produces y^* in the absence of inefficiency and random noise.

We choose a subset of input vectors based on the proximity of the observed output to y^* by taking $(100 \cdot \alpha)\%$ of the original input data (in a probabilistic sense) as

$$L(y^*)_\alpha = \{\boldsymbol{X}_k : \ y_l \leq y_k \leq y_u, \ \int_{y_l}^{y^*} \hat{\pi}(y)dy = \alpha/2, \ \int_{y^*}^{y_u} \hat{\pi}(y)dy = \alpha/2, \ k = 1, \ldots, n\}, \quad \text{(II.5)}$$

where y_l and y_u, respectively, are the lower and the upper bounds for the subset selection satisfying $y_l < y^* < y_u$, and they are uniquely determined in (II.5) given α, y^*, and $\hat{\pi}(y)$. $\hat{\pi}(y)$ is a density estimate of output that can be estimated by a kernel density estimator (Rosenblatt, 1956; Parzen, 1962) with a Gaussian kernel as

$$\hat{\pi}(y) = \frac{1}{nh} \sum_{k=1}^{n} K\left(\frac{y - y_k}{h}\right). \quad \text{(II.6)}$$

$K(\cdot)$ is the standard normal density function, and h is the bandwidth parameter in a kernel estimator. We estimate h based on the direct plug-in method proposed by Sheather and Jones (1991).

Let $(\boldsymbol{X}_i, y_i)$ for $i = 1, \ldots, n_\alpha$ denote the resulting data pair whose input vector is a member of $L(y^*)_\alpha$ where n_α is the number of selected observations. The value of n_α needs to be large enough to estimate the shape of the base isoquant effectively, and it depends on the choice of α. We recommend adjusting α so that $n \cdot \alpha$ is never smaller than $15p(p-1)$. This allows the potential of at least $5(p-1)$ hyperplanes to represent the boundary of a p-dimensional convex set, and each hyperplane in a p-dimensional space is uniquely defined by p points requiring at least p parameters to estimate. For estimating each parameter, we recommend using no less than 3 data points. Such a minimum works well with our Monte

Carlo experiments (see Section II.5).

In (II.5), the symmetric selection in terms of probability guarantees that the sample median of y_i for $i = 1, \ldots, n_\alpha$ is y^* whereas the sample mean becoming y^* requires the symmetry of the sample distribution of y_i's. Let $\bar{y}$ denote the sample mean of y_i's. Also, suppose that $y_i = \phi(\boldsymbol{X}_i) - \epsilon_i'$ where ϵ_i' for $\forall i = 1, \ldots, n_\alpha$ are *i.i.d.* with a mean of $\mu' \geq 0$, i.e., assuming homogeneous errors. Then, considering the randomness coming from the subset selection, thus from the sampling distribution, the (unknown) sample mean of $\phi(\boldsymbol{X}_i)$'s becomes $\bar{y} + \mu'$. By minimizing some random deviation of input vectors $\boldsymbol{X}_i \in L(y^*)_\alpha$, we actually estimate an isoquant associated with $\bar{y}+\mu'$ which we will use as a base isoquant. Since $\bar{y}$ depends on the selected data and μ' is unknown, we refer to the isoquant to be estimated as $\text{Isoq}L(y^*)_{\text{CM}}$ where the subscript CM implies "conditional mean." The usage of $\text{Isoq}L(y^*)_{\text{CM}}$ as a base isoquant can be justified by Remark 1.

Remark 1. *The homotheticity of ϕ implies that $\phi(\boldsymbol{X})$ is invariant to the selection of the base output. In other words, for any y there exists a monotone increasing function f such that $\phi(\boldsymbol{X}) = F\left(g(\boldsymbol{X})\right) = f\left(\theta(\boldsymbol{X}, y)\right)$ where $\theta : \Re_+^{p+1} \to \Re_+$ is the univariate input score aggregating multi-dimensional inputs given a fixed output y. Such a function f can be defined by $f(a) := F\left(a \cdot \kappa(y)\right)$ where $\kappa : \Re_+ \to \Re_+$ is a monotone increasing function of y enabling conversion between the two scaling functions F and f while satisfying $g(\boldsymbol{X}) = \kappa(y) \cdot \theta(\boldsymbol{X}, y)$.*

Remark 1 supports an arbitrary selection of base output associated with a base isoquant if the selection is followed by estimation of a proper scaling function. Given a base output y^*, we estimate $\text{Isoq}L(y^*)_{\text{CM}}$, quantify the corresponding univariate score $\theta(\boldsymbol{X}, y^*)_{\text{CM}}$, and estimate a scaling function f as a function of $\theta(\boldsymbol{X}, y^*)_{\text{CM}}$ in sequence.

Non-uniqueness of $\phi(\boldsymbol{X}_i)$ for $\boldsymbol{X}_i \in L(y^*)_\alpha$ caused by the subset selection in (II.5) induces radial deviation of $\boldsymbol{X}_i$ from the targeting isoquant $\text{Isoq}L(y^*)_{\text{CM}}$, i.e., $\boldsymbol{X}_i = \tilde{\boldsymbol{X}}_i \cdot \epsilon_i$

where $\tilde{\boldsymbol{X}}_i$ is a point on Isoq$L(y^*)_{\text{CM}}$ that shares the same angular coordinates with $\boldsymbol{X}_i$ and $\epsilon_i \in \Re_+$ accounts for the radial deviation. Equivalently, we have $\tilde{\boldsymbol{X}}_i = \gamma_i \boldsymbol{X}_i$ for $\gamma_i = 1/\epsilon_i$. For the estimation of Isoq$L(y^*)_{\text{CM}}$, we estimate $\tilde{\boldsymbol{X}}_i$ by $\hat{\boldsymbol{X}}_i = \hat{\gamma}_i \boldsymbol{X}_i$ for $i = 1, \ldots, n_\alpha$ while minimizing the radial deviation of $\boldsymbol{X}_i$ from the isoquant (equivalently from $\tilde{\boldsymbol{X}}_i$), which can be written as $1 - 1/\gamma_i$. Meanwhile, we restrict the fitted isoquant to be a convex and monotone decreasing function with respect to any input so that the fitted estimate satisfies general characteristics of an isoquant. Consequently, for the estimation of Isoq$L(y^*)_{\text{CM}}$, we solve the following nonlinear programming:

$$
\begin{aligned}
z = \quad & \min_{\hat{\boldsymbol{\gamma}}, \boldsymbol{\alpha}, \boldsymbol{\beta}} \sum_{i=1}^{n_\alpha} (1 - 1/\hat{\gamma}_i)^2 && \text{(II.7a)} \\
\text{s.t.} \quad & \hat{\gamma}_i x_{ip} = \alpha_i + \hat{\gamma}_i \sum_{q=1}^{p-1} \beta_{iq} x_{iq}, \quad \forall i \in \mathcal{I}, && \text{(II.7b)} \\
& \alpha_i + \hat{\gamma}_i \sum_{q=1}^{p-1} \beta_{iq} x_{iq} \geq \alpha_j + \hat{\gamma}_i \sum_{q=1}^{p-1} \beta_{jq} x_{iq}, \quad \forall i, j \in \mathcal{I},\ j \neq i, && \text{(II.7c)} \\
& \beta_{iq} \leq 0, \quad \forall i \in \mathcal{I},\ \forall q = 1, \ldots, p-1, && \text{(II.7d)} \\
& \alpha_i \geq 0, \quad \forall i \in \mathcal{I}, && \text{(II.7e)}
\end{aligned}
$$

where $\boldsymbol{X}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$, and α_i and β_{iq} for $q = 1, \ldots, p-1$ defines a hyperplane containing $\hat{\boldsymbol{X}}_i = (\hat{\gamma}_i x_{i1}, \hat{\gamma}_i x_{i2}, \ldots, \hat{\gamma}_i x_{ip})$, the estimate of $\tilde{\boldsymbol{X}}_i$. The index set $\mathcal{I}$ includes $i = 1, \ldots, n_\alpha$.

The objective function in (II.7a) minimizes the sum of squared radial deviations. If $\phi(\boldsymbol{X}_i)$ for $\forall \boldsymbol{X}_i \in L(y^*)_\alpha$ is constant, $\boldsymbol{X}_i$'s themselves are the points on the relevant isoquant. In this case, $\gamma_i = 1$ for $\forall i \in \mathcal{I}$, and the objective function is zero.

Equations in (II.7b) construct hyperplanes passing through the projected input vector $\hat{\boldsymbol{X}}_i$ for $i \in \mathcal{I}$ by using an arbitrary input as a functional response of the other inputs; here the pth input is chosen. The inequalities in (II.7c), applying the Afriat inequality (Afriat,

1967) to the hyperplanes formed by the projected input vectors, impose the convexity while (II.7d) constrains the monotonicity. Although (II.7e) always holds under other constraints and non-negativity of the observed input vectors, the inequalities are included to improve convergence of the nonlinear optimization.

The task of solving the nonlinear optimization problem in (II.7) can be conducted by using a nonlinear programming solver such as CONOPT available in GAMS. The CONOPT algorithm is based on the generalized reduced gradient (GRG) algorithm (Abadie and Carpentier, 1969) and is made more efficient for large models (Drud, 1985; Drud, 1994). Regardless of which solver is used, nonlinear optimization in general significantly benefits from a good initial solution and reasonable bounds of the decision variables, see Appendix B.1 where we discuss this issue. Appendix B.2 illustrates consistency of the isoquant estimator under multiplicative and radial random noise via Monte Carlo simulation.

II.3.2 Base output selection

Recall from Remark 1, any level of output can serve as a base output y^*. Still, selection of the base output is important for the estimator's finite sample performance. For any given y^*, the subset selection in (II.5) identifies input vectors $\boldsymbol{X}_i \in L(y^*)_\alpha$ that radially deviates from $\text{Isoq}L(y^*)_{\text{CM}}$ but is relatively close to the isoquant. This implies that there is an isoquant associated with the maximum output level among those envelope $\forall \boldsymbol{X}_i \in L(y^*)_\alpha$ from below and another isoquant associated with the minimum output level among those envelope all the input vectors from above. Let us refer to these (unknown) isoquants as a lower bound and a upper bound of our isoquant estimator, respectively. When estimating $\text{Isoq}L(y^*)_{\text{CM}}$ by minimizing the radial deviations of the input vectors, the resulting estimate is expected to pass through the middle of the radial spread of the input vectors. If the radial spread is well distributed around $\text{Isoq}L(y^*)_{\text{CM}}$ for all angular coordinates, the

isoquant estimate should be close to the underlying truth as illustrated in Appendix B.2. Otherwise, the shape of the isoquant estimate can be significantly different from the underlying truth for finite samples, but such difference can be alleviated if the lower and upper bounds get closer to each other. For the base output y^*, we thus prefer an output level that provides the tightest bounds.

The true isoquants are unknown and must be estimated, so derivation of the lower and upper bounds is not possible in actual implementation. Instead, we use some information practically achievable while estimating an isoquant. We consider any y_k for $k = 1, \ldots, n$ for the base output y^*, and then determine the subset of data $(\boldsymbol{X}_i, y_i)$ for $i \in \mathcal{I}$ and solve the optimization problem in (II.7). Let z_k for $k = 1, \ldots, n$ denote the objective function value of the optimization problem associated with y_k, which quantifies the magnitude of the radial deviation of $\boldsymbol{X}_i$'s from the isoquant estimate. In general, the smaller the z_k is, the tighter bounds we can expect. Thus, we prefer a y_k that provides a smaller z_k. Since z_k is calculated as the sum of squared residuals, it tends to increase if the subset of data includes more observations, i.e., if n_α is larger. To eliminate the effect of n_α, we instead use a root mean square deviation (RMSD) type of metric,

$$\text{RMSD}_k = \sqrt{z_k / n_\alpha(k)}, \quad k = 1, \ldots, n,$$

where $n_\alpha(k)$ reveals dependency of n_α on the selection of y^*.

For the selection of the base output, O&R suggest considering three criteria, namely, the selection should yield: 1) the most observations for the isoquant estimation; 2) the largest cone spanned by the isoquant estimate; and 3) the most uniform (angular) spread of the point-wise isoquant estimate. All three criteria can indeed affect the quality of an isoquant estimate. Because the subset selection in (II.5) controls n_α to some extent and we can manage n_α by adjusting α, we only consider the latter two for our base output

selection. Let angle_k and unif_k, respectively, denote the corresponding metrics; see Appendix B.3 for the calculation of the two metrics.

We select the base output based on the three metrics of RMSD_k, angle_k, and unif_k evaluating the tightness of the bounds of the estimate, the range of angles spanned by the estimate, and the difference between the angular distribution of the point-wise estimates and the uniform distribution, respectively. To apply the multiple criteria, we first choose good candidates among all output observations those satisfy all the criteria at a desirable level. Then, we choose an output level of which RMSD_k is the minimum among the candidates.

We consider at least 10% of all output observations for the candidates for the base output. To choose the candidates, we apply a single threshold common to all metrics, so we evaluate the metrics in terms of percentiles instead of their original values. Initially, we use a stringent threshold, e.g., best 10% for each metric (for RMSD_k and unif_k, the smaller the better, whereas for angle_k, the larger the better). If there is any metric which is not within the best 10% range, the corresponding output observation is not selected for the candidate, rendering the number of selected candidates less than 10% of n, the number of all observations. Then, we gradually relax the threshold by iteratively increasing the percentage by a small number until we find at least 10% of the data satisfying the relaxed threshold.

A user may decide, considering one's own application, the percentage of the data selected for the candidates, which also determines the initial threshold. If one prefers equal weights for the three criteria, one may select only one observed output that satisfies the threshold condition at the best available threshold value; this observation then directly becomes the base output. By choosing 10% of data, we assign a higher weight to our first criterion, the tightness of the bounds, while maintaining the other criteria at a sufficiently satisfactory level. In addition, a user may consider an alternative base output selection pro-

cedure based on a single criterion. We present a single criteria alternative in Appendix B.4 which simplifies the overall estimation procedure by using some approximations.

II.3.3 Input aggregation

By applying the methods described in the previous sections, we obtain point-wise estimates of Isoq$L(y^*)_{\text{CM}}$, $\hat{\boldsymbol{X}}_i = (\hat{\gamma}_i x_{i1}, \ldots, \hat{\gamma}_i x_{ip})$ for $i \in \mathcal{I}$. Among the point-wise estimates, only some of them are considered as vertices determining the shape of the isoquant estimate. Let $\boldsymbol{\xi}_l = (\xi_{l1}, \ldots, \xi_{lp})$ for $l = 1, \ldots, n_g$ denote the vertices which can be identified by the quickhull algorithm (see Barber et al. 1996 and `www.qhull.org`). While using these vertices as benchmark, we calculate $\theta(\boldsymbol{X}_k, y^*)_{\text{CM}}$ as a norm ratio of an input vector $\boldsymbol{X}_k$ to its projection onto the Isoq$L(y^*)_{\text{CM}}$ estimate. The optimization problem for solving for the norm ratio is given below:

$$
\begin{aligned}
\min_{\tau_k,\, \boldsymbol{s},\, \boldsymbol{\lambda}} \quad & \tau_k \\
\text{s.t.} \quad & \tau_k x_{kq} - \sum_{l=1}^{n_g} \lambda_l \xi_{lq} - s_q = 0,\ \forall q = 1,\ \ldots,\ p, \\
& \sum_{l=1}^{n_g} \lambda_l = 1, \\
& \lambda_l \geq 0,\ \forall l = 1,\ \ldots,\ n_g, \\
& s_q \geq 0,\ \forall q = 1,\ \ldots,\ p,
\end{aligned}
\tag{II.8}
$$

where τ_k is the objective function value of the optimization problem for $\boldsymbol{X}_k$, and s_q for $q = 1, \ldots, p$ denotes a slack variable for the qth input. λ_l for $l = 1, \ldots, n_g$ is used to describe the convex combination of $\boldsymbol{\xi}_l$. Using the solutions of (II.8), the aggregate input estimate is calculated as $\hat{\theta}(\boldsymbol{X}_k, y^*)_{\text{CM}} = (\tau_k)^{-1}$. Similar to O&R, we remove observations with $s_q > 0$ for any q, for the subsequent analysis.

II.4 Estimation of a production function consistent with the regular ultra-passum law

Given aggregate input estimate $\hat{\theta}(\boldsymbol{X}_k, y^*)_{\text{CM}}$, we estimate the scaling function f specified in Remark 1 to complete the estimation of the homothetic production function ϕ. For simplicity, we refer to aggregate input $\hat{\theta}(\boldsymbol{X}_k, y^*)_{\text{CM}}$ 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 (II.1) and Remark 1, we have

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

With subtraction and addition of mean inefficiency μ in the right hand side,

$$y = [f(x) - \mu] + [\mu - u + v] = \psi(x) + e, \tag{II.9}$$

where $\psi(x) := f(x) - \mu$ is called an average-practice production function and $e := \mu - u + v$ denotes an error term with $E(e) = 0$. We estimate f by estimating ψ and then shifting ψ upwards with the estimate of the mean inefficiency μ, similar to Aigner et al. (1977) and Kuosmanen and Kortelainen (2012).

For univariate or aggregate input, an S-shaped function such as ψ 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 (Banker et al., 1984) and Free Disposal Hull (FDH, Deprins et al. 1984) to find efficient observations. Then, they 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 overestimate the production function as they envelope all data points from above; the magnitude of the overestimation increases as the noise level increases.

We now present a new procedure to estimate a production function imposing the RUP law while accounting for noise in the stochastic setting. In what follows, we present a method to estimate the average-practice curve in Section II.4.1, to estimate the mean inefficiency in Section II.4.2, and to estimate the production frontier in Section II.4.3.

II.4.1 Average-practice curve estimation using CNLS

To estimate an S-shaped curve, we generate a fine grid over the domain of the univariate (or aggregate) input and refer to it by x_g for $g = 1, \ldots, m$ where m is the number of grid points. We consider each grid point as a potential inflection point location with respect to the input. Given x_g, we divide the input domain into two regions where the functional estimate is convex and concave, respectively. We use CNLS to estimate the convex and concave parts of the S-shaped curve. Our estimates of the inflection point location and the S-shaped curve then correspond to those providing the smallest sum of squared errors (SSE) from the CNLS fitting.

Specifically, for each x_g, $g = 1, \ldots, m$, we partition the input data into two subsets: $\{x_k : x_k < x_g,\, k = 1, \ldots, n\}$ defining the convex region and $\{x_k : x_k > x_g,\, k = 1, \ldots, n\}$ defining the concave region. Then, we rearrange the data set $\{(x_k, y_k) : k = 1, \ldots, n\}$ in non-decreasing order of the input values, so that the observations in the convex region are indexed as from 1 to n_{cvx} and those in the concave region from $n_{cvx} + 1$ to n. Simply applying CNLS to each region and integrating the estimates from both regions does not guarantee the estimates meet at the inflection point. Thus, we solve the CNLS problems

for both regions in a unified optimization problem:

$$\min_{\boldsymbol{\beta}} \quad z_g = \quad \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \tag{II.10a}$$

$$\text{s.t.} \quad \hat{y}_k = \beta_{k,0} + \beta_{k,1} x_k, \ \forall k = 1, \ldots, n, \tag{II.10b}$$

$$\beta_{k,1} = \frac{\hat{y}_{k+1} - \hat{y}_k}{x_{k+1} - x_k}, \ \forall k = 1, \ldots, n_{cvx}, \tag{II.10c}$$

$$\beta_{k,1} = \frac{\hat{y}_k - \hat{y}_{k-1}}{x_k - x_{k-1}}, \ \forall k = n_{cvx} + 1, \ldots, n, \tag{II.10d}$$

$$\beta_{k,1} \leq \beta_{k+1,1}, \ \forall k = 1, \ldots, n_{cvx} - 1, \tag{II.10e}$$

$$\beta_{k-1,1} \geq \beta_{k,1}, \ \forall k = n_{cvx} + 2, \ldots, n, \tag{II.10f}$$

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

$$\beta_{k,0} \leq 0, \ \forall k = 1, \ldots, n_{cvx}, \tag{II.10h}$$

$$\hat{y}_k \geq 0, \ \forall k = 1, \ldots, n. \tag{II.10i}$$

The constraints in (II.10c) – (II.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 (II.10b), (II.10c), (II.10e), and (II.10g) are equivalent to those of CNLS fitting a monotone increasing and convex function with a single input, whereas (II.10b), (II.10d), (II.10f), and (II.10g) substitute the constraints of CNLS modeling for the concave part.

From (II.10c) and (II.10d), we have $\beta_{n_{cvx},1} = (\hat{y}_{n_{cvx}+1} - \hat{y}_{n_{cvx}})/(x_{n_{cvx}+1} - x_{n_{cvx}})$ and $\beta_{n_{cvx}+1,1} = (\hat{y}_{n_{cvx}+1} - \hat{y}_{n_{cvx}})/(x_{n_{cvx}+1} - x_{n_{cvx}})$ representing the slopes of the last and the first hyperplane in the convex and the concave region, respectively. These constraints define a hyperplane connecting the convex and the concave region. In the presence of (II.10e)–(II.10f), therefore, the S-shaped function estimate obtained from (II.10) is continuous at the inflection point and satisfies the shape constraints.

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 dimension. To be consistent with this requirement, we impose that our production function goes through the origin. It can be proven that including (II.10h) and (II.10i) ensures the production function estimate goes through the origin (see Proposition 4 in Appendix B.5 for the statement of sufficiency and proof).

After solving (II.10) for each inflection point candidate x_g, we estimate the inflection point, denoted by x^*, as x_g that minimizes the objective function value of (II.10) among all x_g, i.e., $\hat{x}^* = x_g$ such that $z_g = \min\{z_g : g = 1, \ldots, m\}$. Similarly, we estimate $\psi(x_k)$ as $\hat{y}_k$ for $k = 1, \ldots, n$ obtained from the optimization problem with the minimum objective function value. Under some regularity conditions, we can prove the consistency of the estimator as shown in Theorem 1.

Theorem 1. *Let the average practice production function, $\psi(x)$, be differentiable up to the second order. For given n, let $x^*_- = max\{x_k : x_k < x^*, k = 1, \ldots, n\}$ and $x^*_+ = min\{x_k : x_k > x^*, k = 1, \ldots, n\}$, and suppose that the sequences of $\{x^*_-\}_n$ and $\{x^*_+\}_n$ both approach x^* as $n \to \infty$. Also, let the number of grid points tend to infinity in the way that there is at least one grid point between x^*_- and x^*_+ while $n \to \infty$. Then, as $n \to \infty$ and $m \to \infty$, $\hat{x}^* \to x^*$. Furthermore, the corresponding least square estimator (with the minimum objective) is a consistent estimator of $\psi(x)$.*

Theorem 1 requires infinite number of grid points, i.e., $m \to \infty$, in order to obtain the desired results, but practically a large number is typically sufficient.

II.4.2 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 μ_k instead of a constant μ, so that the model in (II.9) is expressed as,

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

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 clustering strategy. The strategy starts with partitioning the univariate input domain into a number of disjoint intervals (called bins) that will be used to cluster observations with similar input values. Such an action is known as "binning," which is frequently used to generate a histogram (a discrete estimate of probability density) in statistics. We let each cluster thereby each bin involve a uniform number of observations, n_b. Subsequently, we assume that the composite errors are homoscedastic or identically distributed within a cluster.

This strategy is similar to the usage of the k-Nearest Neighbor (k-NN) method in Simar (2007) in the sense that the fixed number of observations form a neighborhood or a cluster; but it is different from the k-NN method because our strategy does not allow overlap of the intervals (bins). Simar (2007) apply the k-NN method for constructing a neighborhood that ensures enough data points when estimating stochastic DEA/FDH efficiency along a ray in an input-oriented case. After projecting those data in the neighborhood onto the ray, they apply the method described in Hall and Simar (2002) to deconvolude the inefficiency and noise terms.

The choice of n_b is such that it creates as many clusters as possible but should maintain enough observations in each cluster. A good choice, in our experience, is to base n_b on the result of the average-practice function estimation. Specifically, we let n_b be the largest integer such that $n_b \leq n/m_h$, where m_h is the number of hyperplanes in the average-practice function estimate. In other words,

$$n_b = \max\{a : a \leq n/m_h,\, a \in \mathbb{Z}_+\}, \tag{II.12}$$

where $\mathbb{Z}_+$ is a set of positive integers. If n/m_h is less than a minimum threshold (set as 7 observations per cluster), we recommend to set n_b to the minimum threshold value; the experiments in Simar (2007) illustrate that the optimal choice for the number of observations to use is no less than 7 observations. For the same reason, if there are fewer than 7 observations in the last bin, we merge the last bin into the nearest preceding bin.

Depending on the applications, one may use cross validation (CV) to choose the optimal n_b. In our application of the wind power curve estimation, however, CV does not provide much benefit compared to the selection in (II.12). In general, the optimal solution from CNLS fitting results in a relatively small number of hyperplanes, m_h, regardless of how large the number of observations is. For the analysis in Section II.6, we have $n \approx 13,000$ and $m_h \approx 40$, so $n_b \approx 325$. This means that we have a sufficient number of data points in each cluster while imposing a similar level of heteroscedasticity observed in the average practice curve. On the other hand, calculating CV scores for each candidate of n_b ranging from 1 to approximately 13,000 would require substantial amount of calculation which is not likely to be feasible.

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

assume for each cluster

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

where $u_{b,\,o}$ and $v_{b,\,o}$ are *i.i.d.* within the b th cluster. Due to the *i.i.d.* assumption, $E(u_{b,\,o}) = \mu_b$ for $\forall o$, meaning that mean inefficiency is constant within each cluster. Since $u_{b,\,o}$ is a non-negative random variable, the density of $(\mu_b - u_{b,\,o})$ has jump discontinuity at μ_b, and the density is zero if $(\mu_b - u_{b,\,o})$ is greater than μ_b. With this variable specification, we apply the method described in Hall and Simar (2002) to estimate the cluster-wise mean inefficiency μ_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.

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. In other words, the jump discontinuity of the density of $e_{b,\,o}$ is induced by that of $(\mu_b - u_{b,\,o})$ if the variance of $v_{b,\,o}$ is quite small. Following Hall and Simar's method, we estimate μ_b by finding a point where the first derivative of the density of $e_{b,\,o}$ is maximized in absolute terms. Algorithm 1 explains how to apply Hall and Simar's method for each cluster.

Algorithm 1 Estimating cluster-wise mean inefficiency

Require: h_{HS}: bandwidth parameter, Δh_{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 $\forall o = 1,\ \ldots,\ n_b$ applying kernel density estimation with bandwidth h_{HS}.

Compute $\tilde{\mu}_b$ by finding a positive root of $\hat{p}''(e)$ in a neighborhood of $\hat{\mu}_b$ where $\hat{p}(e)$ is the density estimate of the residuals
if $|\tilde{\mu}_b - \hat{\mu}_b| \leq T$ **then**
$h_{HS} \leftarrow h_{HS} - \Delta h_{HS}$
$T \leftarrow |\tilde{\mu}_b - \hat{\mu}_b|$
$\hat{\mu}_b \leftarrow \tilde{\mu}_b$
end if
until $|\tilde{\mu}_b - \hat{\mu}_b| > T$
return $\hat{\mu}_b$

We set the initial value of the bandwidth parameter h_{HS}, suggested to be a large number 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 3σ 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 Δh_{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 in the shape of the density estimate.

II.4.3 Production frontier curve estimation

Using the relationship $f(x_k) = \psi(x_k) + \mu_k$, we can construct a frontier function estimate, $\tilde{f}$, employing cluster-wise mean inefficiency estimates, i.e., $\tilde{f}(x_k) = \hat{\psi}(x_k) + \sum_{b=1}^{m_h} \hat{\mu}_b \mathbb{1}_{I_b}(x_k)$, where I_b is a set of input values in the bth cluster. The values of $\tilde{f}$, shifted upward from $\hat{\psi}$ with different constants, no longer satisfy the RUP law or the shape constraints. A similar problem occurs in Simar (2007) where a frontier estimator collecting individual DEA/FDH efficiency estimators for different rays does not preserve

usual properties of DEA/FDH estimators such as monotonicity or convexity. Their strategy is to smooth the individual estimators based on kernel smoothing and refit the frontier by running DEA or FDH program. In our case, we simply refit the shifted cluster-wise production function segments $\tilde{f}$ by using the extended CNLS in (II.10) in order to obtain a continuous frontier function estimate satisfying the shape constraints. Note here that in this refitting process, we do not fit to data, but we do fit to the estimated values $\tilde{f}(x_k),\ k = 1, \ldots, n$ that already have noise and inefficiency removed.

To use the extended CNLS in (II.10), we again involve the grid points for the potential inflection point location, similarly to what has been done for the average-practice curve estimation in Section II.4.1. If summarizing, for each $x_g,\ g = 1,\ \ldots,\ m$, we divide the input domain into two parts defining the convex region and the concave region. Instead of using the original data pair (x_k, y_k), however, we now use $(x_k, \tilde{f}(x_k))$ for $k = 1,\ \ldots,\ n$ to fit the curve using (II.10). This provides an estimate of the frontier curve that satisfies the shape constraints and the continuity. Our estimate $\hat{f}$ then corresponds to the one with the minimum SSE among m different estimates determined by each grid point.

II.5 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 several variants.

The O&R DGP generates 100 observations with two inputs and one output. For two-dimensional input, the O&R DGP 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 according to $y = \phi(x_1, x_2) = F\left(g(x_1, x_2)\right)$, 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 σ_v varies in magnitude. Specifically, for a given $\tilde{\sigma}_u$, we consider five different σ_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 σ_u/σ_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),

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(\phi(\boldsymbol{X}_k) - \hat{\phi}(\boldsymbol{X}_k) \right)^2}. \tag{II.13}$$

For the simulation study in this section and wherever applies, we set the user-defined parameters as follows. When selecting the base output as in Section II.3.2, we consider at least 10% of all output observations for the base output candidates and set the initial threshold of multiple criteria to 10%. To select the subset of data estimating a base isoquant as in (II.5), we set $\alpha = 0.05$. However, for the results in this section where $n = 100$, α is restricted by the minimum threshold for $n \cdot \alpha$ of $15p(p-1) = 30$, so α becomes

0.3. For the estimation of average-practice curve and production frontier curve, we use 100 grid points, i.e., $m = 100$, to split the (univariate) input space into the convex and the concave regions. Finally, to select the cluster size n_b, we apply (II.12).

Figure II.2 demonstrates 90% confidence 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 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 II.3 compares O&R's aggregate input estimator, the true univariate input $g(\boldsymbol{X}_k)$, and HJD's aggregate input estimator, all applied with HJD's scaling function estimator used in the second stage. 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, using HJD's aggregate input estimator performs significantly better than using O&R's aggregate input; this is shown in Figure II.4. This comparison

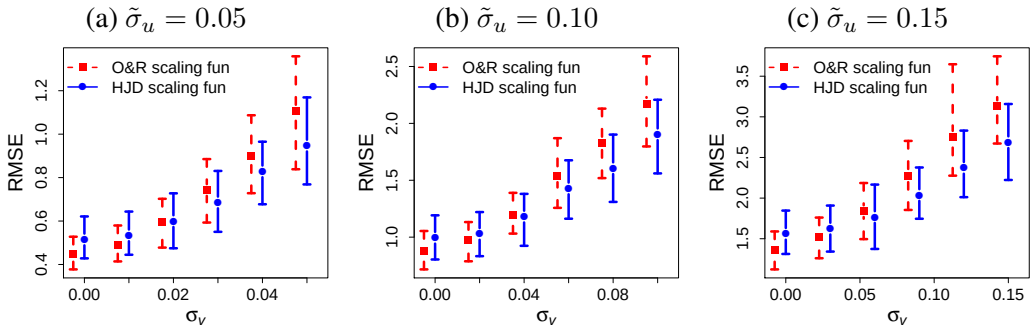

Figure II.2: Result of scaling function estimation based on true univariate input.

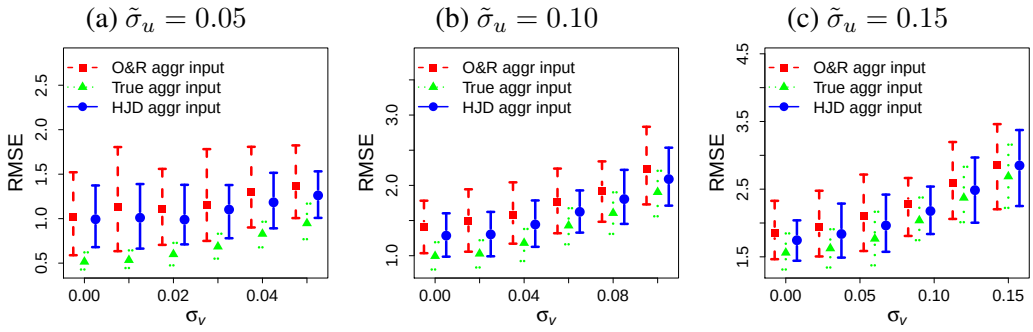

Figure II.3: Comparison of the aggregate input estimations using HJD's scaling function estimator in the second stage.

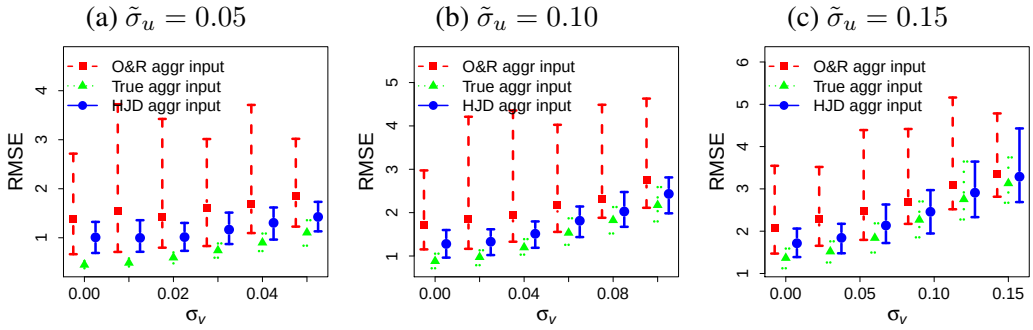

Figure II.4: Comparison of the aggregate input estimations using O&R's scaling function estimator in the second stage.

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

Finally, Figure II.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 pronounced.

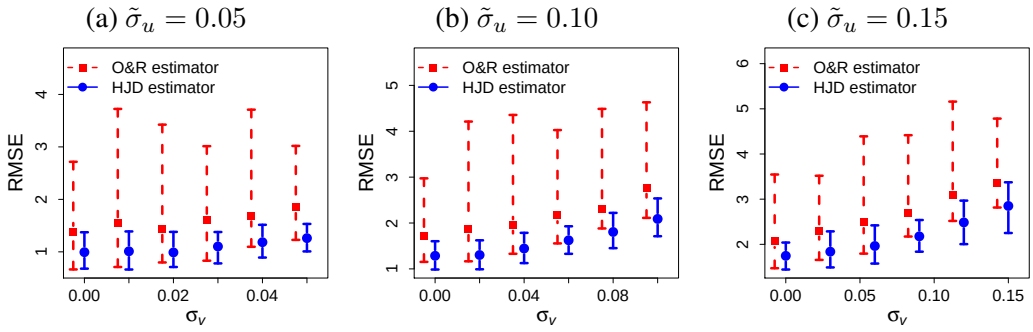

Figure II.5: Comparison of O&R and HJD production function estimators.

II.6 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 II.6(a) shows a typical nominal power curve provided by the turbine's manufacturer that 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 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 the wind speed-versus-power, one can

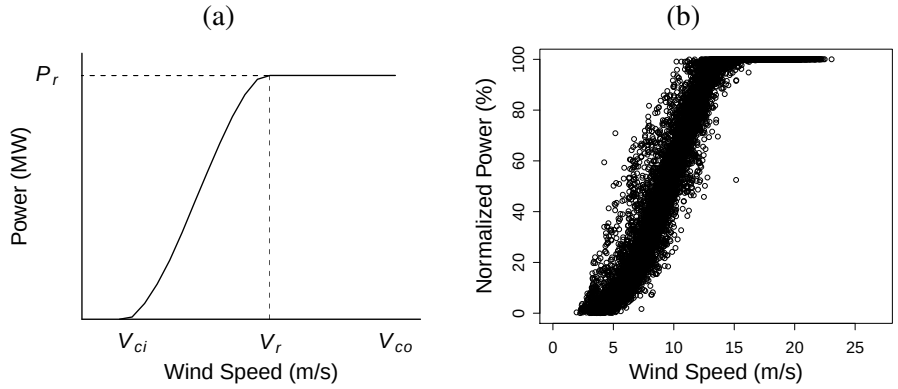

Figure II.6: Wind turbine data: (a) nominal power curve and (b) scatter plot of the data used in this example.

observe that the actual measurements scatter broadly around the nominal curve; see Figure II.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. Fitting the model in (II.9) provides a reference power curve indicating the best practice performance of a wind turbine in the presence of noise. Such a reference curve is critical when evaluating efficiency of a wind turbine. In addition, given some knowledge on how the efficiency evolves over time, it is possible to analyze actual investment recovery of a wind project during its life cycle. Common practice uses a nominal power curve as a reference curve; but in fact, a nominal power curve is just an average practice curve obtained under a controlled experimental setting and may differ from the wind turbine's power output in an actual operating environment.

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 II.6 and II.7, we

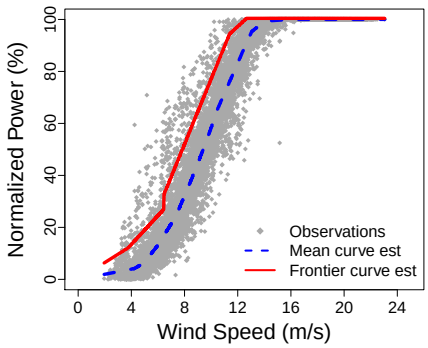

Figure II.7: Power curve estimation: production function estimates. No data are measured in the range of wind speed from 0 m/s to approximately 2 m/s. Note that both the average-practice function estimate and frontier function estimate extend to (0,0) in this range.

normalize the power response (vertical axis) by setting the rated power to be 100%.

We have wind speed and power data measured in 2007. The wind industry typically measures wind speed and power data as the average of 10-minute observation intervals. We have approximately 13,000 observation pairs of wind speed and power output after pre-processing (eliminating missing data, etc.) for one particular wind turbine. 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 II.7 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 0.895 that implies 10.5% of systematic

inefficiency on average.

II.6.1 Monte Carlo simulation for wind turbine application

To investigate suitability of the proposing estimator for wind power curve application, we simulate a set of data similar to the one for the wind turbine and compare with the deterministic estimator based on the simulated dataset. We revise O&R's DGP that is used in Section II.5 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 listed below:

(i) 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 $F(z) = 15/(1 + \exp(-3.5 \ln z))$.

(ii) 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 σ_{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 $\sigma_{vk} = \tilde{\sigma}_v$.

(iii) 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 ϕ, 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 present in the system. The mean and standard deviation of the composite error in the real data are 1.89 and 1.54, respectively, after rescaling for similarity to the O&R's DGP.

Figure II.8(a) shows the scatter plot of a set of simulated data. If we choose $\tilde{\sigma}_u = 2.2$ and $\tilde{\sigma}_v = 1.32$ 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 II.8(b). The dotted line corresponds to $\tilde{\sigma}_v = 1.32$, indicating where the characteristics of our wind turbine data would fall in this analysis. Figure II.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 also better suited for estimating future power curves.

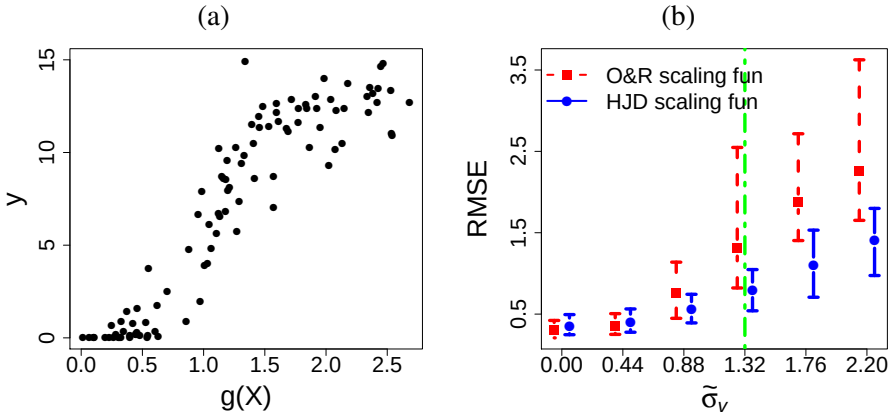

Figure II.8: Simulation results: (a) scatter plot of simulated data mimicking wind turbine data and (b) comparison results using the simulated data with various scales of noise. For (b), $\tilde{\sigma}_u = 2.2$ with varying $\tilde{\sigma}_v$ which is indicated along the x-axis.

II.7 Conclusions

This chapter presents a new RUP law-satisfying estimator, which also allows heteroscedasticity in both the random noise and inefficiency modeling. 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 requiring the use of a stochastic estimator. Recall, 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 increase, resulting in even poorer performance of the deterministic estimators.

Finally, we feel that future research is needed to relax some assumptions we made. The use of the input homotheticity 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 Färe and Shephard (1977). Developing the ray-homothetic structure into a stochastic estimator represents a considerable improvement. Another aspect to consider is that the proposed estimator examines the production function with a single output only, similar to the one in Olesen and Ruggiero (2014). Further study addressing how to handle multiple outputs is useful as multi-output production functions are of growing importance in the production function literature.

CHAPTER III

A PRODUCTION ECONOMICS ANALYSIS FOR QUANTIFYING THE EFFICIENCY OF WIND TURBINES*

We quantify the productive efficiency of a wind turbine, using power output and environmental variable data, measured either at the turbine or at a meteorological mast near the turbine. The methods described can potentially help with decision makings in asset procurement, maintenance planning, or wind turbine control optimization. The current recommendation from the International Electrotechnical Commission regarding turbine performance evaluation is to use a power curve or power coefficient. What is commonly used in practice is the average performance power curve or power coefficient. When using the power curve to quantify productive efficiency, one crucial shortcoming is the lack of a common best performance benchmark while the power coefficient approach uses an absolute efficiency measure that is not achievable. We apply the new approach developed in Chapter II for efficiency quantification based upon production economics' concepts that provides estimates of a best performance benchmark. Our specific approach has two main components: (a) a best performance power curve is estimated and used together with the average performance curve to show how well a turbine has performed relative to its full potential; and (b) a covariate matching procedure is developed to control for environmental influences for the comparison of turbine performances over different periods. Through a simulation study, we demonstrate that the proposed efficiency is more sensitive to potential changes in the turbine. When analyzing multi-year wind turbine data, we observe that the turbine's efficiency is improving during the first two years of operation and then

*Reprinted with permission from Hwangbo, H., Johnson, A. L., and Ding Y. (2017). A production economics analysis for quantifying the efficiency of wind turbines. *Wind Energy* (in press) `http://dx.doi.org/10.1002/we.2105`, Copyright © 2017 by John Wiley & Sons, Ltd.

remains relatively constant during years 3 and 4.

III.1 Introduction

Among important issues to be addressed for making wind energy more competitive, one concerns performance quantification of wind turbines. Addressing this issue adequately, namely, quantifying a turbine's productive efficiency and understanding its change over time, helps guide numerous decisions for operating wind turbines; for instance, planning maintenance actions to counter degradation in turbine performance (Márquez et al., 2012), justifying costly retrofitting turbine upgrades (Lee et al., 2015a), or optimizing pitch and torque control (Abdullah et al., 2012) to prolong a turbine's service life. Performance quantification enables performance benchmarking of turbines from different manufacturers, which could also help with decision making in the asset procurement process.

As discussed in Section I.2, the performance evaluation metrics recommended by International Electrotechnical Commission (IEC) are not sufficient to represent productive efficiency of wind turbines, and some efficiency metrics derived from them have significant drawbacks for quantification of the productive efficiency. The limitations of the existing metrics motivate us to look into the field of production economics (Hackman, 2008). In production economics, efficiency quantification is based on the estimation of a production function and the explicit modeling of systematic inefficiency, using input and output data for a set of production units, be it firms, factories, hospitals or power plants. In the context of wind energy, a wind turbine is a power production unit, wind speed is the dominating input driving power production, and the generated power is the output.

Recently, some production function estimation methods from production economics have been used in wind energy applications. Two studies (Carvalho et al., 2009; Pieralli et al., 2015) are noted but neither of them adequately addresses two important characteristics of wind turbine data, namely, (i) the wind turbine data are inevitably contaminated

by noises and (ii) the data appears to follow an S-shape, as shown in Figure I.4, comprising a convex region, followed by a concave region, and the two segments of curves are connected at an inflection point.

Carvalho et al. (2009) simply applied the DEA approach to the wind-power data, which envelopes all observations from above with a piece-wise concave function. Pieralli et al. (2015) applied a different approach, known as free disposal hall (FDH). FDH relaxes the concave function assumption but still assumes noise-free observations. Production economics researchers call this type of frontier analysis approach, assuming noise-free observations, *deterministic*. The problem with applying a deterministic approach to noisy wind production data is that it tends to overestimate the best performance benchmark because every observation is assumed to be achievable. Figure III.1(a) and III.1(b) show, respectively, the production frontiers estimated using the DEA and FDH approaches.

A recently developed DEA-type estimator explicitly imposing the Regular Ultra Passum (RUP) law (Olesen and Ruggiero, 2014) may be a better choice than the two methods described above since this estimator at least imposes the S-shape of the curve revealed in the wind data. However, the frontier analysis employed by Olesen and Ruggiero (2014)

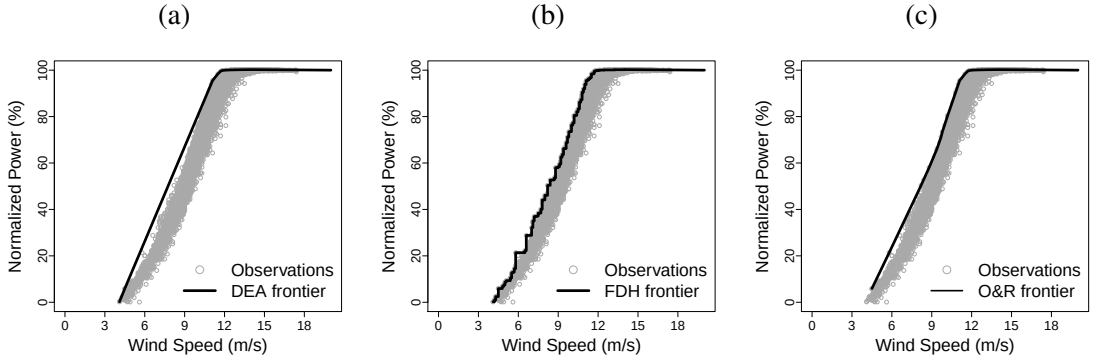

Figure III.1: Various types of production function: (a) data envelopment analysis, (b) free disposal hull, and (c) deterministic S-shaped production function.

is still deterministic, enveloping all observations from above, and consequently suffers from the overestimation that all other deterministic production function estimators suffered. Figure III.1(c) presents an example of the production frontier estimated by the Olesen-Ruggiero (O&R) method, when it is applied to a set of wind turbine data.

In this chapter, we discuss a production economics approach that addresses the aforementioned wind data characteristics, namely, significant noise and the particular shape of the functional relationship revealed in observed data, and hence is more suitable for estimating a production function of a wind turbine. We use the estimated production function as a performance benchmark for a wind turbine, which will be used together with the average performance power curve to quantify productive efficiency of the turbine.

Although wind speed is the dominating force driving in wind power production, other environmental variables including wind direction, humidity, and turbulence intensity as well as air density (as shown in the power production equation, (I.1)) may all affect wind power output (Lee et al., 2015b). Many of the environmental variables are measured at the meteorological mast. In order for the resulting productive efficiency measure to be used for performance comparison between different turbines or for the same turbine over different time periods, it is important to control for the environmental influences, so that the performance comparison quantifies the differences coming from a turbine's endogenous characteristics. For this reason, we develop a covariate matching procedure, allowing us to select a subset of the data, for which the probability distributions of the environmental variables are matched. Note that controlling for environmental influences is also necessary when either AEP, power curve, or power coefficient is used as a performance measure. Because of this, the covariate matching procedure is potentially widely applicable.

The rest of this chapter is organized as follows. Section III.2 discusses how to apply the ideas from production economics to the efficiency quantification of wind turbines. Section III.3 establishes a method to control for environmental influences by matching

the probability densities of the environmental covariates. Section III.4 presents a case study applying our proposed efficiency measure to actual wind turbine data. Section III.5 concludes the chapter.

III.2 Estimation of performance benchmark and efficiency quantification

We use a shape constrained functional estimation to model the typical shape of wind data, which is consistent with the RUP law, and model noise. The estimator is described in detail in Chapter II; we summarize the key features below.

The basic production function model for a wind turbine can be expressed as

$$y = f(V) - u(V) + \epsilon, \tag{III.1}$$

where y refers to the power output produced by a turbine and V is wind speed. To be consistent with the IEC standards (IEC12.1, 2005), we denote by V the wind speed normalized by air density, i.e., $V = V_{10min} \cdot (\rho_{10min}/\rho_0)^{1/3}$ where V_{10min} and ρ_{10min} are, respectively, wind speed and air density averaged over 10-minute time intervals, and ρ_0 is the average of the measured air density at the test site during the periods of data collection. Hereinafter in this chapter, we refer to this normalized wind speed as "wind speed" unless otherwise stated. Random noise ϵ is assumed having a zero mean, while the systematic inefficiency term $u(V)$ is a non-negative random variable with positive mean, i.e., $\mu(V) := E[u(V)] > 0$. Note that $u(V)$ is a function of V, meaning that the amount of inefficiency varies as the input changes, which is referred to as a heteroskedastic inefficiency term.

The basic production function model in (III.1) is re-written as:

$$y = [f(V) - \mu(V)] + [\mu(V) - u(V) + \epsilon] = g(V) + e, \tag{III.2}$$

if we let $g(V) := f(V) - \mu(V)$ and $e := \mu(V) - u(V) + \epsilon$. The expression connects the power curve with the production function. To see this, consider the following. The term e is a redefinition of the error term with expectation zero. Because of this, $g(V)$ is the average practice power curve. As such, the production function $f(V)$ differs from the power curve $g(V)$ by the mean of the inefficiency varying by V.

This connection helps lay out the intuition behind the procedure of estimating $f(V)$. One would start with a power curve from the wind turbine data, then estimate the mean function of the inefficiency and use it to rotate the average power curve to the new position resulting in the production frontier function.

We stress that because the final $f(V)$ needs to satisfy the RUP law (i.e., the S-shape constraint), the average performance power curve $g(V)$ that comes before the production function must satisfy the same shape constraint. This requirement makes our power curve estimation procedure different from those currently used in practice because none of them imposes the S-shape constraint explicitly. Common practice including the standard procedure recommended by IEC (Lee et al., 2015a; IEC12.1, 2005) estimates a power curve nonparametrically because the functional form of the power production equation is unknown. The resulting estimates, however, still tend to approximate an S-shape curve, but with noticeable local differences from a strictly S-shaped curve. We will present an example below in Figure III.2.

Estimating the shape constrained power curve $g(V)$ requires imposing convexity and concavity in the low and high wind speed regions, respectively. The convex segment should connect to the concave segment at the inflection point, which itself needs to be estimated from the data. The estimation of the convex segment or the concave segment can be done by using the method Convex Nonparametric Least Squares (CNLS) (Kuosmanen, 2008). When the two segments are estimated simultaneously maintaining the shape constraints and the continuity at the inflection point, the final outcome of this step is the

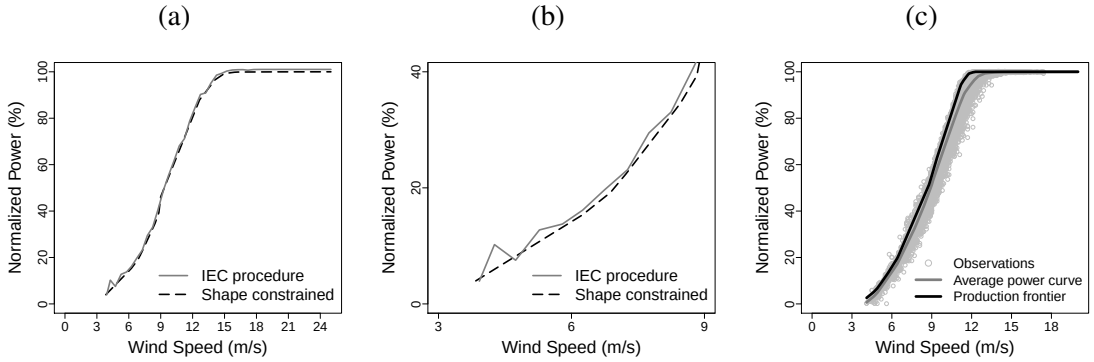

Figure III.2: Illustration of stochastic S-shaped production function: (a) comparison to the IEC standard procedure, (b) comparison to the IEC standard for the wind speed ranging from 3 m/s to 9 m/s, (c) estimates of average practice power curve and production frontier.

average performance power curve $g(V)$.

After $g(V)$ is estimated, we can take differences between the fitted power curve and the output y. According to the relationship in (III.2), the resulting residuals are the summation of two random components: $\mu - u$ and ϵ. Our modeling assumption states that u is non-negative and ϵ is symmetrically distributed with respect to a zero mean. So we expect to see a significant decrease in the density of the residuals at the value of μ. This understanding is used to estimate μ, which is unknown. If we can locate where the greatest decrease in the residual distribution occurs, it gives us an estimate of μ. Specifically, the technique in Hall and Simar (2002) can be used for this estimation.

The following summarizes the steps in estimating the shape constrained stochastic production function $f(V)$:

(i) Use the wind turbine data (wind speed and power) to estimate $g(V)$ while imposing the shape constraints and the continuity requirement at the inflection point; denote the estimated curve by $\hat{g}(V)$.

(ii) Estimate $\mu(V)$, the mean function of the inefficiency term;

(iii) Estimate $f(V)$ based on the relationship of $f(V) = g(V) + \mu(V)$; denote the estimated curve by $\hat{f}(V)$.

The steps of the procedure have been simply described here, and for technical details of the estimation procedure, see Chapter II.

Figure III.2(a) presents two average performance power curves: one satisfies the shape constraint, obtained by the procedure outlined above, whereas the other obtained by the IEC's standard procedure does not. The two estimates are similar to each other but not the same (see the enlarged version in Figure III.2(b)). Figure III.2(c) presents the production frontier curve and the average performance curve, both satisfying the shape constraint. Compared with the deterministic estimators shown in Figure III.1, one notices that this production frontier does not envelop all the observations. The observations beyond the frontier are affected by significant positive random noise.

With the average performance power curve and the best performance benchmark estimated, we propose the following efficiency measure, θ, which is the ratio of the energy produced under the average performance power curve over that under the best performance, integrated over the whole wind spectrum:

$$\theta = \frac{\int_{V_{ci}}^{V_{co}} \hat{g}(V)dV}{\int_{V_{ci}}^{V_{co}} \hat{f}(V)dV}, \tag{III.3}$$

where V_{ci} is the cut-in wind speed and V_{co} is the cut-out wind speed. Apparently, θ takes a value between 0 and 1; the closer θ is to 1, the closer the wind turbine performs to its full potential.

III.3 Controlling for environmental influences through covariate matching

We estimate both the best practice frontier curve and average performance power curve as a function of wind speed. Besides wind speed, air density and several other environ-

mental variables, including wind direction, humidity, turbulence intensity and wind shear, all potentially affect the wind power production. These environmental influences are not controllable, but their existence does play a role affecting the inefficiency estimated from the power output data. Consequently, when comparing the productive efficiency of different turbines or the same turbine over different operational periods, practitioners often wonder what part of inefficiency is due to the turbine's intrinsic differences and what part of inefficiency comes from differences in environmental characteristics such as air dampness. This sort of ambiguity can be alleviated if the comparison periods have comparable environmental profiles. Creating comparable environmental profiles is what we try to accomplish in this section.

Let us consider monitoring a turbine's efficiency change over a number of time periods. The environmental variables are referred to as covariates in statistics. The covariate vector includes measurements from the wind farm as well as those computable using available measurements (such as wind shear), but it does not include variables shown in previous studies to have little or no correlation to power output. We acknowledge that wind farms may gather different data on environmental measurements; for instance, one of the wind farms we worked with does not have humidity measurements. Nonetheless, our procedure presented here can be applied regardless of the number of variables included in the covariate vector.

We describe a method to match covariate vectors to make the environmental profiles across different time periods as similar as possible, thus removing the effect of environmental influences from the efficiency analysis. Suppose that we have p environmental variables in the covariate vector and t periods of operation. We assume that for different periods, the common p variables are available. If not, we reduce the set of variables included in the covariate vector to the subset common to all periods. We arbitrarily choose one of the periods as the reference period.

Let $\boldsymbol{X}^t = (x_1^t, \ldots, x_p^t)$ for $t = 1, \ldots, T$ denote the vector of covariates consisting of p environmental variables, including wind speed and air density, observed during the tth period and y^t be the corresponding power output. The reference period is denoted by $t' \in \{1, \ldots, T\}$. We recommend setting $t' = T$ so that the analysis is based on the most recent data. A non-reference period, also called an evaluation period, is denoted by $t \neq t'$. The data pairs in periods t' and t are represented, respectively, by $(\boldsymbol{X}_j^{t'}, y_j^{t'})$ for $j = 1, \ldots, n_{t'}$ and $(\boldsymbol{X}_k^t, y_k^t)$ for $k = 1, \ldots, n_t$, where $n_{t'}$ and n_t are the number of observations in the two periods, respectively.

Our matching procedure starts with selecting a single observation in the reference period, comparing its environmental covariates (i.e., variables in $\boldsymbol{X}$) with the covariates of an observation in an evaluation period, and assessing their dissimilarity using a score defined below. Repeat this for all observations in the evaluation period and select the observation yielding the smallest dissimilarity score as the best match. If no observation in the evaluation period has a small enough score, the observation is removed from the reference period set. Otherwise, choose another evaluation period and find the best match to the observation in the reference period. As such, for a single observation in the reference period, there will be one matched observation from each of the evaluation periods. Altogether this produces a set of matched covariate vectors having similar environmental profiles. We then proceed with the same action for all observations in the reference period. Figure III.3 illustrates this procedure.

Now let us define the dissimilarity score used in the matching process. Consider the j-th observation in the reference period t'. For the q-th variable in the covariate vector, we denote by $\mathcal{S}_{kq}$, $q \in \{1, \ldots, p\}$ and $k \in \{1, \ldots, n_t\}$, the dissimilarity score between this reference observation and the k-th observation in the evaluation period t. The dissimilarity

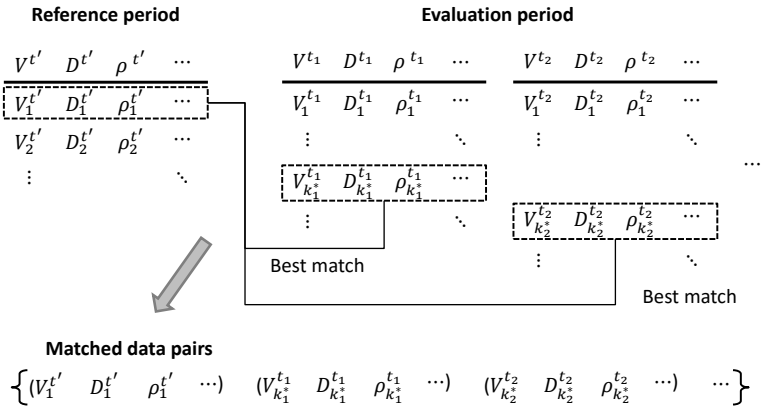

Figure III.3: Procedure to construct a set of matched covariate vectors.

score, $\mathcal{S}_{kq}$, is defined as:

$$\mathcal{S}_{kq} = \frac{|x_{jq}^{t'} - x_{kq}^{t}|/sd_q^{t'}}{x_{jq}^{t'}/mean_q^{t'}}. \tag{III.4}$$

The smaller the dissimilarity score is, more similar the two covariate vectors are. In the above definition, $mean_q^{t'}$ and $sd_q^{t'}$ are, respectively, the sample mean and the sample standard deviation of the q-th variable in period t', and their use is to normalize the scale of the preceding terms. A normalization is needed because two covariates can have different ranges both in an absolute value sense and in a percentage sense; for instance, air density varies by 15 percent from its mean value, while wind speed can vary up to 100 percent. Without the normalization, covariates with low variability will be labeled as matching, even though their density functions differ significantly.

The above formula can be used for almost any environmental variables, except for wind direction, which is a circular variable for which the value 0 and 360 are equivalent. For wind direction, we slightly modify the dissimilarity score as:

$$\mathcal{S}_{kq} = \frac{\min\left\{|x_{jq}^{t'} - x_{kq}^{t}|,\, 360 - |x_{jq}^{t'} - x_{kq}^{t}|\right\}}{x_{jq}^{t'}} \cdot \frac{mean_q^{t'}}{sd_q^{t'}}. \tag{III.5}$$

Once $\mathcal{S}_{kq}$ is calculated, we find the set of candidate best matches to the j-th observation in t' that satisfies $\mathcal{S}_{kq} \leq \omega$, where ω is a pre-specified threshold and usually is set to a small quantity, for example, 0.25. The use of ω is to set a standard for eligible matches, such that any resulting matches are considered "good enough." If there is no observation in the evaluation period satisfying this dissimilarity constraint, then this j-th observation in period t' is skipped as having no matched record. On the other hand, if there are multiple observations in this candidate set, we choose the best match, indexed as k^* that satisfies the following minimax criterion:

$$k^* = \underset{k \in \mathcal{K}}{\operatorname{argmin}} \left\{ \max_{q \in \{1, \ldots, p\}} \mathcal{S}_{kq} \right\}, \tag{III.6}$$

where $\mathcal{K} = \{k : \mathcal{S}_{kq} \leq \omega, \forall q = 1, \ldots, p\}$. Without the ω threshold and using the minimax criterion alone, one could end up with a match whose dissimilarity score may be uncomfortably large.

Once the matching process is done for all evaluation periods, for notational simplicity, we re-index the matched data pairs by i, such that $(\boldsymbol{X}_i^t, y_i^t)$, $i = 1, \ldots, n$ and for $t = 1, \ldots, T$ where n is the number of the matched data pairs.

Note that the above matching process does not produce an exact match but a good match, subject to the dissimilarity allowed by the threshold ω. To confirm the quality of the matches, we suggest plotting the probability density functions (pdf) of each environmental variable, empirically estimated from the data and visually inspected to assess how well the pdfs match across the comparison periods. Numerical examples will be presented in the case study section to further illustrate this point.

To evaluate the productive efficiency of a wind turbine controlling for the environmental influences, we use the matched data pairs, namely $\{(\boldsymbol{X}_i^t, y_i^t)\}$, to estimate the average practice power curve $g(V)$ and the best performance benchmark $f(V)$. Without prior

knowledge of the period when a wind turbine shows the best performance, the best practice curve cannot be restricted to a specific t. For this reason, we pool the matched data pairs from all periods (including t') while estimating $f(V)$. In order to see how the turbine productive efficiency may have changed from period to period, we further estimate the average practice power curve for each t. As such, the efficiency measure in Equation (III.3) can be re-expressed as:

$$\theta_t = \frac{\int_{V_{ci}}^{V_{co}} \hat{g}_t(V)dV}{\int_{V_{ci}}^{V_{co}} \hat{f}(V)dV}, \tag{III.7}$$

where $\hat{g}_t$ is the average practice curve of period t.

III.4 Case study

In this case study, we use data from two onshore wind turbines (WT1 and WT2) and two offshore wind turbines (WT3 and WT4). Table III.1 summarizes the characteristics of these wind turbines; for certain entries an approximation rather than the accurate value is given for the protection of the identities of the turbine manufacturers and wind farms. The wind turbine data include observations during the first four years of their operations. In particular, the onshore data are available for the period of 2008–2011, and the offshore data are from 2007 to 2010. The four years worth of data allows us to look into the performance change during the early stage of a turbine's operation. All measurements are the averages during 10-minute time intervals, a common practice for data arrangement in the wind industry.

We analyze the wind turbine data on an annual basis. In other words, we divide the four-year data into four consecutive annual periods, namely that we have $T = 4$ and $t = 1, 2, 3, 4$. We evaluate turbine efficiency for each year because seasonal variations in atmospheric and meteorological conditions are significant, but yearly patterns are relatively stable.

The measurements we have for the offshore wind farm include power output (y), wind

speed (V), wind direction (D), air density (ρ), turbulence intensity (I), wind shear (S), and humidity (H). For the onshore wind farm, however, the humidity measurements are not available. The power output is always measured on the wind turbine. Most of the environmental measurements are taken from a meteorological mast closest to the turbine, with the exception of wind speed and turbulence intensity which are measured on the wind turbine. The mast measurements are used either because some variables are only measured at the mast (such as air pressure and ambient temperature, which are used to calculate air density) or because the mast measurements are considered more reliable (such as wind direction). The wind speed measurements are taken from the nacelle anemometers, and they are further used to calculate the turbulence intensity. The industrial partners who provided the data told us that the wind speed data, measured by the nacelle anemometer, have been adjusted to be the free stream equivalents in front of a turbine, rather than the raw measurements taken after the wake of a turbine's rotor. We use the nacelle anemometer measurements for wind speed to better differentiate the wind turbines nearby the same mast.

Prior to analyzing the data, we conducted some preprocessing, removing data records with missing values or data records taken while a turbine is unavailable, or excluding measurements such as negative power values. These observations seem to occur randomly,

Table III.1: Specification of the wind turbines.

	Onshore	Offshore
Location	U.S.	Europe
Cut-in wind speed (m/s)	3.5	3.5
Cut-out wind speed (m/s)	20	25
Rated wind speed (m/s)	approximately 13	approximately 15
Rated power (MW)	1.5–2	approximately 3
Initiation of operations (year)	2008	2007

so we do not anticipate a sample selection bias.

After the preprocessing, we select the subset of data with comparable environmental profiles through the covariate matching method described in Section III.3. For onshore wind turbines, the covariates to be matched include $\boldsymbol{X} = (V, D, \rho, I, S)$, whereas for offshore wind turbines $\boldsymbol{X} = (V, D, \rho, H, I)$. The reason that we did not include wind shear S for offshore wind turbines is because a previous study (Lee et al., 2015b) found that conditioned on the inclusion of (V, D, ρ, H, I), the effect of S on power output appears negligible. We did test to see what happened if we included wind shear in the offshore turbine data matching process. It turns out that the results of that analysis produced the same insights described below.

For all turbine cases, we use $\omega = 0.25$ as the threshold assessing the dissimilarity. Before the covariate matching, the number of observations in each annual dataset ranges from 14,000 to 37,000, and these numbers reduce to 1,400–2,300 after the matching. The significant reduction in the number of observations indicates the importance of matching covariates. Had we used all the raw observations, the efficiency results would describe the differences in the operating environment across periods, rather than the intrinsic efficiency of the turbine. The matched data set still includes thousands of observations that is a large enough sample for estimating the best performance benchmark as well as the average performance curve.

Figure III.4 and III.5 present the pdfs of each environmental variable across the four comparison periods after the covariate matching; Figure III.4 is for onshore turbine WT1, while Figure III.5 is for offshore turbine WT3. We omit the plots for WT2 and WT4, which are similar, in the interest of space. Notice that the choice of $\omega = 0.25$ leads to sufficiently good matching as demonstrated in the pdf plots.

Subsequently, we use the matched subset of data to estimate the productive efficiency measure for each comparison period, as defined in Equation (III.7). Because of the ran-

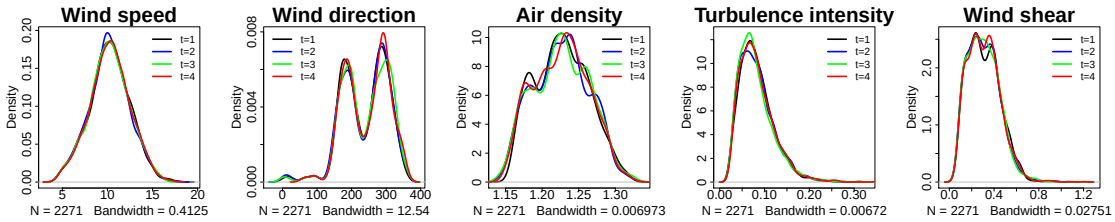

Figure III.4: Probability density function plots of the matched covariates over the four comparison periods for onshore turbine WT1.

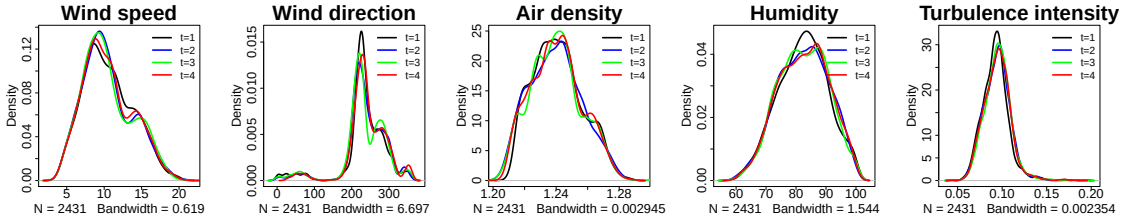

Figure III.5: Probability density function plots of the matched covariates over the four comparison periods for offshore turbine WT3.

domness in the data, we add a confidence interval to the efficiency measure. To do that, we use a bootstrap procedure, which is to resample the data with replacement, B times, and for each resampled dataset, compute the efficiency measure, which altogether results a total of B replications. Then, the confidence interval for the efficiency measure can be constructed using these B sample replications; for details about the bootstrap procedure, please refer to Hastie et al. (2009). In this case study, we used $B = 100$ and calculate 90% confidence intervals.

Figure III.6 shows the productive efficiency θ_t and its confidence intervals for the four comparison periods, which are the first four years of a turbine's operation. Interestingly, we note that for all four turbines, their productive efficiency appears to have increased slightly, rather than deteriorated, during the early stage of operation. This pattern is more

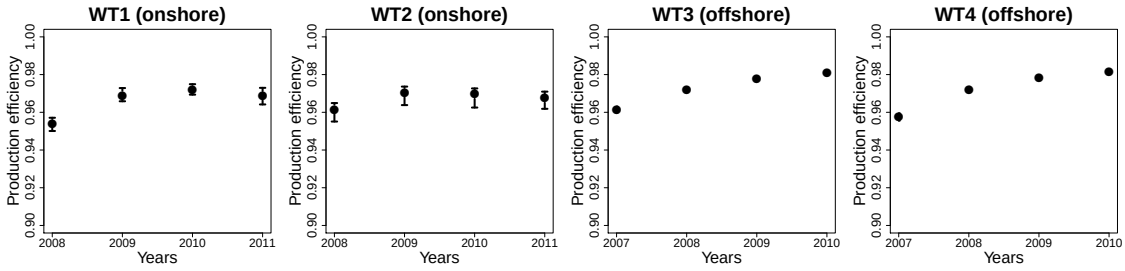

Figure III.6: Productive efficiency θ_t, $t = 1, 2, 3, 4$. The bars represent 90% confidence intervals and the dots denote the mean values of the efficiency. For offshore wind turbines, the confidence intervals are very narrow, so that the bars are not shown explicitly.

obvious for offshore turbines. This initial increase in efficiency was also recognized by Staffell and Green (2014). Figure 9b in Staffell and Green (2014) plots the fleet-level performance degradation of wind turbines over a twenty-year period using the fleet's load factor as the performance measure. Staffell and Green (2014)'s study appears to suggest an initial period of four to five years before any noticeable degradation was witnessed, as well as an increase in turbine performance for the first one-and-half years, which is quite consistent with what we observed.

Next, we want to compare the proposed productive efficiency measure and power coefficient, given the popularity of power coefficient used in turbine performance evaluation. We calculate the peak power coefficient values for each comparison period, using the same matched subset of data and the power coefficient curves averaged for each yearly period. We also apply the bootstrap procedure to compute the 90% confidence intervals of the (peak) power coefficient. The power coefficient values and the proposed productive efficiency values are presented in Table III.2, in which the values in the parenthesis are the respective confidence intervals.

As we mentioned before, the power coefficient itself is not a relative measure. One could divide a power coefficient by the Betz limit to get a similar interpretation as the

Table III.2: Comparison between the productive efficiency θ_t and the (peak) power coefficient: the values represent the mean of the bootstrap estimates and the values in parenthesis are the respective 90% confidence intervals.

	Power coefficient				Productive efficiency θ_t			
	Year 1	Year 2	Year 3	Year 4	Year 1	Year 2	Year 3	Year 4
WT1	0.371	0.388	0.393	0.393	0.954	0.969	0.972	0.969
	(0.367, 0.377)	(0.386, 0.392)	(0.390, 0.397)	(0.389, 0.398)	(0.950, 0.957)	(0.966, 0.973)	(0.969, 0.975)	(0.964, 0.973)
WT2	0.444	0.466	0.463	0.462	0.962	0.970	0.970	0.968
	(0.439, 0.450)	(0.461, 0.470)	(0.460, 0.468)	(0.457, 0.467)	(0.955, 0.965)	(0.964, 0.974)	(0.963, 0.973)	(0.962, 0.971)
WT3	0.420	0.465	0.483	0.505	0.962	0.972	0.978	0.981
	(0.417, 0.423)	(0.461, 0.473)	(0.479, 0.488)	(0.497, 0.511)	(0.960, 0.963)	(0.971, 0.973)	(0.977, 0.979)	(0.980, 0.982)
WT4	0.417	0.473	0.484	0.506	0.958	0.972	0.978	0.981
	(0.417, 0.421)	(0.465, 0.484)	(0.477, 0.494)	(0.496, 0.516)	(0.956, 0.959)	(0.971, 0.973)	(0.977, 0.980)	(0.980, 0.983)

productive efficiency value. Given that the yearly power coefficient ranges from 0.371 to 0.506, the relative power coefficient efficiency would be between 63% to 85%. Please bear in mind that the Betz limit is a theoretical limit impractical to attain. So these low percentages should be taken into account with perspective; they should not be interpreted as saying that power production of the wind turbines is inefficient. If we look at the productive efficiency values, the wind turbine operations are actually reasonably efficient, relative to their full potentials.

Using the power coefficient values, we also notice a general upward trend and a leveling off. This message appears to reinforce what we found using the productive efficiency measure. In fact, there appears a fairly obvious positive correlation between the two measures; using all the values in Table III.2 yields a correlation of 0.70 between power coefficient and the proposed productive efficiency. This positive correlation suggests that the proposed productive efficiency measures a turbine's performance on a broad common ground with the power coefficient.

One may wonder what is then the benefit of using the proposed productive efficiency measure rather than the power coefficient. To address this, we present a study below based on Khalfallah and Koliub (2007), in which they investigate how dust accumulation on

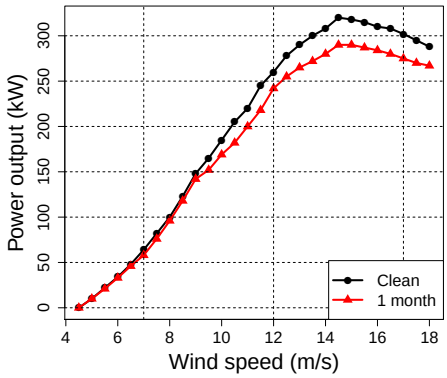

Figure III.7: Effect of dust accumulation on turbine blades: redrawn referring to Figure 2 in Khalfallah and Koliub (2007).

turbine blades affects turbine performance. They analyze wind turbines operated in Egypt where the air at the turbine site is very dusty. In Figure III.7, we regenerate one of their graphs that compare the power production performance of a wind turbine when the blades are clean versus when they are exposed to dust accumulation for a month. Please note that with the dust accumulation, the power performance deteriorates more significantly for wind speed higher than 9 m/s than the lower wind speeds. Khalfallah and Koliub (2007) also compare the average power loss for a stall-regulated turbine and for a pitch-regulated and concluded that a pitch-regulated turbine suffers a smaller loss, around 3%.

Without the actual production data from the Egyptian turbines, we create a set of simulated data, mimicking the dust accumulation effect. We take the WT1 data measured during 2008 and modify it by decreasing the power output value by 3% (because our turbine is pitch regulated) for those power outputs corresponding to wind speed of 9 m/s or higher. We treat the resulting data as for 2009. Then, we reduce the 2008 power data by 6% and 9% and use them as the substitute of 2010 and 2011 power data, respectively. All environmental data are left intact.

Figure III.8 illustrates the change of the power coefficient (left panel) and the produc-

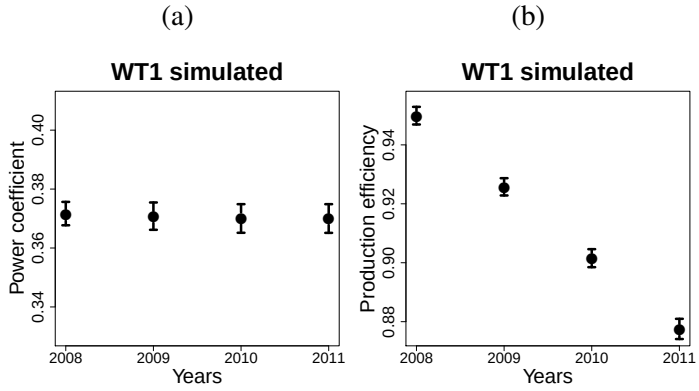

Figure III.8: Change of wind turbine efficiency implied by (a) power coefficient and (b) productive efficiency. The bars represent 90% confidence intervals and the dots denote the mean values of the corresponding efficiency measures.

tive efficiency (right panel) over the four-year period. Again, we use 100 bootstrap replications to compute the 90% confidence intervals, which show up as the bars in the plots. The use of the productive efficiency shows a clear trend, signifying the dust accumulation effect over the year. The average year-to-year decrease in the productive efficiency is about 2.4%. This magnitude of decrease seems reasonable as the 3% power reduction initially imposed is only applied to a subset of data (wind speed higher than 9 m/s). By contrast, the power coefficient does not show any trend in the change of power production ability. Recall that the power coefficient is the peak point representation of the power coefficient curve (the point C in Figure I.1(c)) and, as such, using it could miss an underlying change that does not happen to the peak point area. For the original 2008 data, the peak value on the power coefficient curve is obtained at the wind speed ranging from 7.5 m/s to 8.5 m/s, whereas the dust accumulation affects power production beyond that wind speed region. We acknowledge that the power coefficient metric could possibly reveal better than shown in this analysis if a larger range of wind speeds, instead of only the peak value, is used. The challenge is, of course, to find a proper method that aggregates the power efficient

curve covering a large range of wind speeds for the purpose of characterizing a turbine's efficiency.

III.5 Concluding remarks

Wind turbine operators often wonder how efficiently their turbine generators have been producing power relative to a practically attainable optimal case. For this purpose and taking advantage of ideas and methods in production economics, we introduce a method to estimate the best achievable performance in the operation of wind turbines. Determining such a benchmark provides a reference for defining a normalized measure, thus quantifying the productive efficiency of a wind turbine. Compared with the current industrial practices, the proposed productive efficiency measure involves both the best performance benchmark and the average performance curve, whereas the power curves or power coefficients are average performance measure. Our case study shows that the proposed productive efficiency is more sensitive to a change in a turbine's production capability. When applied to the first four years of data on two pairs of turbines, we observe an increasing pattern in terms of productive efficiency in the initial operation of a turbine. This observation corroborates the findings in an independent study.

CHAPTER IV

SPLINE MODEL FOR WAKE EFFECT ANALYSIS: CHARACTERISTICS AND IMPACTS ON WIND TURBINE POWER GENERATION

Understanding and quantifying wake effect plays an important role in improving wind turbine designs and operations as well as wind farm layout planning. The majority of the current wake effect models are physics-based, but these models have a number of shortcomings. Sophisticated models based on computational fluid dynamics suffer from computational limitations and are impractical for modeling commercial sized wind farms, whereas simplified physical models are generally inaccurate for wake effect quantification. Nowadays, data-driven wake effect models are gaining popularity as the data from commercially operating wind turbines become available, but this development is still in its early stage. This study contributes to the general category of the data-driven wake effect modeling that makes use of the actual wind turbine operational data. We propose a wake effect model based on splines with physical constraints incorporated, which sets out to estimate wake effect characteristics such as wake width and wake depth. As such, our model is one of the first data-driven models that provide a detailed account of the wake effect. Prediction accuracy of the proposed spline model, when compared with other alternatives, also confirms the benefit of incorporating the physical constraints in the statistical estimation.

IV.1 Introduction

While a wind turbine is operating, the turbine absorbs kinetic energy in wind and converts the energy into electricity. Consequently, the wind behind the rotor loses some of its original kinetic energy, causing reduction in its speed. Moreover, the rotating blades

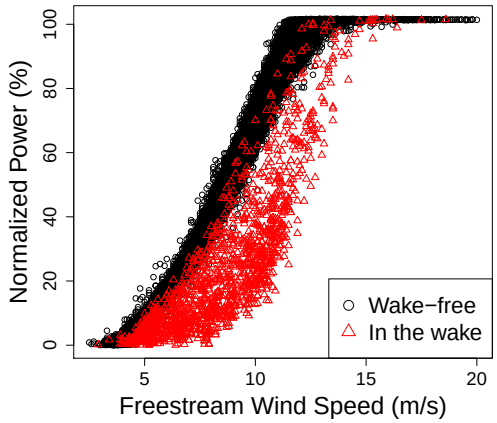

Figure IV.1: Power output in the wake versus that under a free stream condition. free stream wind speed refers to wind speed measurements not affected by another turbine. Due to the confidentiality concern, power output is denoted throughout this chapter by a percentage value normalized to its maximum.

disturb the natural flow of the air and create turbulence for the downstream turbines. Such a phenomenon differentiating the after-rotor wind flow from the free stream one before the rotor is referred to as the wake effect. As the amount of power output depends on ambient wind speed, the reduction in wind speed due to wake may substantially deteriorate power production at downstream wind turbines. Figure IV.1 illustrates power outputs of a wind turbine when it is wake free versus when it is in the wake of another turbine.

There have been significant efforts devoted to a better understanding of wind turbine wake and alleviation of its effect on power generation. Except for a few recent studies that are data-driven (which we will review in the next section), the majority of the current wake effect models are physics-based. A simplified parametric model, yet widely used in practice, is Jensen's model (Jensen, 1983). A primary shortcoming of the Jensen's model lies in its unsatisfactory accuracy in predicting turbine power loss under commercial operating conditions. The limitation of this simplified model has motivated researchers to resort to sophisticated, computational fluid dynamics (CFD) models that can achieve a

higher accuracy (Laan et al., 2015). Yet, using the CFD models faces significant computational challenges. For example, running large eddy simulations, one of the popular CFD methods, requires days or even weeks of computation on supercomputers to analyze a single-wake situation (Sanderse et al., 2011).

In this chapter, we propose a data-driven alternative to the physics-based wake effect models. To facilitate a successful transition from physics-based models to data-driven modeling, we incorporate certain physical understanding and considerations as constraints in the data model fitting procedure. We consider the single wake situations arising between two turbines of which modeling assumptions are easier to justify. Single wake situations are of great interest in the wake studying literature (Prospathopoulos et al., 2011; Duckworth and Barthelmie, 2008), and the single wake behaviors provide valuable insights into various decisions for improving wind farm performance (more details in Section IV.2).

We do want to note an additional difference between the physics-based models and the data-driven models in general. All the physics-based models do not directly estimate the power loss. Instead, they primarily focus on estimating the reduced wind speed due to wake. To quantify wake power loss, these models then require an additional layer of converting the wind speed estimates into a corresponding power output; such a conversion can be done by using a simple power curve model as recommended by the International Electrotechnical Commission (IEC12.1, 2005) or more complicated power curve models recently developed (Lee et al., 2015b). By contrast, the data-driven models, including the one proposed in this chapter, connect the wind data directly to the power output in a single step. Figure IV.2 illustrates the difference.

In the remaining parts of this chapter, we describe in Section IV.2 the wake characteristics with practical implications for improving power generation performance of wind energy systems. We also explain the existing data-driven models in terms of their capability of deriving the wake characteristics. We present in Section IV.3 the proposed wake

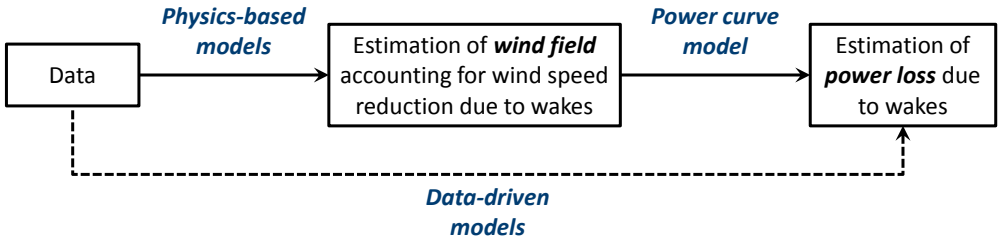

Figure IV.2: Wake power loss estimation procedures.

model based on a spline model structure incorporating physical constraints and the model estimation procedure. We compare the prediction performance of our wake model with that of other alternatives in Section IV.4, and apply it to the analysis of wind turbine wake and its effect on power generation in Section IV.5. In Section IV.6, we conclude the chapter.

IV.2 Characteristics of wake effect and data-driven approaches

Physics-based wake models are able to conform with the physical phenomena and capable of characterizing wake effect, through wake width and depth, for instance. To understand these characteristics, let us consider the setting in Figure IV.3(a). Suppose that there are two turbines, and let θ denote an acute angle between wind direction and the line segment connecting the two turbines. For a given wind direction, the wake caused by Turbine 1 affects the downstream region with a range of angles (the shaded area). Wind speed loss is expected to be the highest along the center line, the line parallel to the wind direction that passes through the upstream turbine (Turbine 1). Then, Turbine 2 is subject to the largest power loss when $\theta = 0$. As θ increases, the wake region rotates accordingly. As such, Turbine 2 deviates from the center line, and the amount of its power loss decreases. After θ exceeds a certain value, Turbine 2 is no longer in the wake of Turbine 1. The maximum power loss when $\theta = 0$ is referred to as the wake depth, whereas

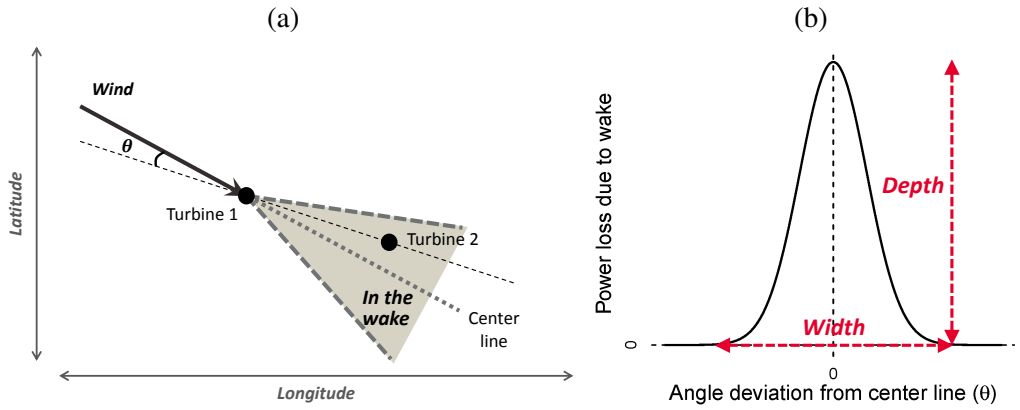

Figure IV.3: Characteristics of wind turbine wake and its effect: (a) wake region and θ; (b) wake depth and wake width. Wake power loss is expected to be a function of θ.

the range of θ for which a turbine is in the wake of another turbine and thus its power loss is positive is referred to as the wake width (see Figure IV.3(b)). Wake depth and width are supposed to remain constant when the relative positions between two turbines are fixed and their values will change when the turbine relative positions are different.

Knowledge of wake characteristics is crucial for improving power generation performance on wind farms. As wake width and depth strongly depend on the relative positions of turbines, characterizing the turbine specific wake effect facilitates the layout planning for wind farms (Khalid and Savkin, 2012; Emami and Noghreh, 2010). Understanding the wake characteristics also supports effective operational control of wind turbines through pitch and yaw controls (McKay et al., 2013; Gebraad et al., 2016). The pitch control can regulate the magnitude of wind speed loss in a downstream region by adjusting the energy absorption level of an upstream turbine, and the yaw control can change the path of wake diffusion, for instance, by tilting the wake region, so that Turbine 2 can be as nearly wake free as possible for a given wind direction.

A common data-driven practice in industry to understand the wake effect is as follows.

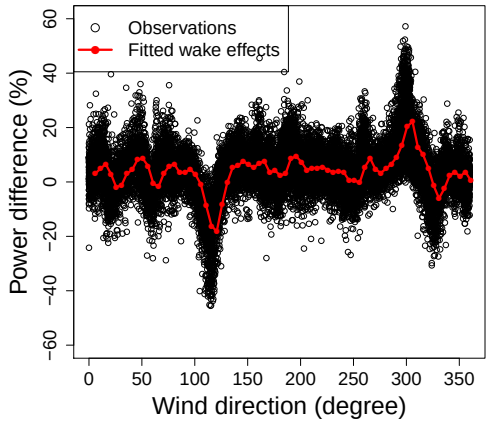

Figure IV.4: Estimation of wake effects between a pair of turbines. The between-turbine distance is four times the rotor diameter d. The distance from this pair to other turbines is more than $10d$.

First, gather the power output data from the two turbines. Second, choose a specific range of wind speed where the maximum power loss is expected, e.g., 8.0 ± 0.5 m/s (Barthelmie et al., 2010). Some other studies extend the coverage of wind speed to a wider range, e.g., 5.0–11.0 m/s (McKay et al., 2013). Then, plot the power difference between the two turbines under the above-specified wind speeds against the wind direction (0 degree means the north). Figure IV.4 shows a plot between the power difference and wind direction. To smooth out the noise effect, people would apply the action of binning, namely, partition the wind direction by a unit, say $5°$, and then, average all the power data points in a specific bin and use the average as the representative of the original data. Applying the data binning to the raw data in Figure IV.4 produces the solid line passing through the data cloud.

The solid line is treated as the estimated curve representing the wake effect. The wake depth can be read from the plot by observing the two peaks around $120°$ and $300°$, respectively. As we move with the wind direction from 0 to 360 degrees, the roles of the two turbines, namely that which one is wake free and which one is in the wake, are reversed. That is why we observe that one of the peaks is upside down. The wake width is not

immediately obvious, so that researchers usually impose a large enough angle coverage, say, $\theta \in (-25^\circ, 25^\circ)$ and then verify with the estimated curve that the angle range is indeed broad enough (McKay et al., 2013). Sometimes one calculates the angle coverage on each side of the center line at which the power loss estimate is within a certain level, for example, $\pm 5\%$ of the free stream power (Barthelmie et al., 2010). When using this data binning method, a purely data-driven method, the estimate of the power difference from the wake-free turbine to the in-wake turbine is not guaranteed to be positive. As a matter of fact, previous studies (Prospathopoulos et al., 2011; Troldborg et al., 2011) often show that some of the bin-wise estimate of this power difference is negative even after θ moves beyond the obvious wake width region; this phenomenon is in fact evident in Figure IV.4.

Other than the data binning approach, there are few other data-driven wake models. One exception is a sophisticated statistical model recently developed (You et al., 2017), which is a Gaussian Markov random field (GMRF) model with a Bayesian hierarchical structure that accounts for spatial correlation of power output of turbines at different locations. To be clear, this GMRF model was not particularly developed for studying the single wake situations. You et al. (2017) makes use of the spatial correlations among turbines located close to one another and simultaneously estimates the heterogeneous power outputs from multiple turbines by modeling the wake interactions with GMRF. Their model is useful for analyzing wake effects in mid- to large-size wind farms, but it can lose estimation accuracy for two turbine settings. Moreover, their model does not impose the constraint that the wake power loss is positive. Nevertheless, due to its consideration of turbines in their relative spatial locations, the GMRF model could estimate the wake power loss indirectly by taking the difference of the maximum fitted value among all turbines and the power output fitted to a specific turbine.

In this chapter, we use the data binning approach and the GMRF model in You et al. (2017) as our benchmarks. You et al. (2017) also review a number of other possible data-

driven alternatives, such as global model (GLB), individual model (IND), and random effect model (RND), but conclude that those other alternatives are not competitive with the GMRF model. Hence, we do not consider those other alternatives any further in this study.

IV.3 Spline-based wake effect model and its estimation

IV.3.1 Baseline power production model

In this study, we borrow a model structure from Chapter II for representing wind power production from the field of production economics, which reads as:

$$y_t(x) = f(x) - \eta_t(x) - \omega_t(x) + \epsilon_t, \quad t = 1, \ldots, T \tag{IV.1}$$

where y represents the power output, x is a vector of predictors that include wind related (and possibly other environmental) variables, and T is the number of turbines. In this model, $f(\cdot)$ represents the best achievable power output characterizing the full power production potential of a specific type of wind turbines given wind resources, also known as the production frontier in the production economics literature (Aigner et al., 1977), $\eta_t(\cdot)$ and $\omega_t(\cdot)$ are the inefficiency terms, namely the power losses relative to the best achievable production, and ϵ_t is the i.i.d. random noise. In Chapter II, $\eta_t(\cdot)$ and $\omega_t(\cdot)$ are lumped into one single inefficiency term. Here, we split them into two, so that $\eta_t(\cdot)$ represents a turbine's inherent inefficiency independent of wake, whereas $\omega_t(\cdot)$ represents the turbine's power loss due to wake. In this model, we need both power loss terms to be non-negative, i.e., $\eta_t(\cdot) \geq 0$ and $\omega_t(\cdot) \geq 0$ for $\forall t = 1, \ldots, T$, to be consistent with the physical understanding of the phenomenon.

The existence of $f(\cdot)$ may sound strange to some readers, but including $f(\cdot)$ does not require any restrictive model assumptions. For a pair of turbines, one can pool the two

turbine power production data together and estimate a common production frontier; the full detail of such estimation process is described in Chapter II. For our study of the single wake situations, however, this $f(\cdot)$ does not even need to be explicitly estimated. As one will see shortly, we will establish a power difference model in the subsequent section, which takes the power difference between a pair of turbines. By doing so, the common frontier disappears in the resulting model.

IV.3.2 Power difference model for two-turbine cases

We now consider the two-turbine case in a single wake situation. We here introduce two angle variables, θ_1 and θ_2, to be associated with the two turbines. Specifically, θ_1 denotes the wind direction causing power loss on Turbine 1 and θ_2 denotes the wind direction under which Turbine 2 endures power loss. As illustrated in Figure IV.5, θ_1 and θ_2 can take any direction in the sets $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively, given the definition of these sets stated below. For the purpose of analyzing the wake effect, θ_1 and θ_2 only need to vary in the 180° outer hemisphere surrounding their respective turbine. Note that θ_1 is actually on the side of Turbine 2, whereas θ_2 is on the side of Turbine 1. If we position the zero degree of θ_1 and θ_2 at the line connecting the two turbines, then $\theta_1, \theta_2 \in (-90^\circ, 90^\circ)$. We denote by $\mathcal{D}_1$ the set of wind directions corresponding to the support of θ_1, and likewise, by $\mathcal{D}_2$ the set of directions in which θ_2 is defined.

With this notation and by following the baseline power production model in (IV.1), we can write the individual power production functions for the two turbines, respectively:

$$\begin{aligned} y_1(x) &= f(x) - \eta_1(x) - \omega_1(x) \cdot \mathbb{1}_{\mathcal{D}_1}(x) + \epsilon_1, \\ y_2(x) &= f(x) - \eta_2(x) - \omega_2(x) \cdot \mathbb{1}_{\mathcal{D}_2}(x) + \epsilon_2, \end{aligned} \tag{IV.2}$$

where $\mathbb{1}_{\mathcal{D}_t}(x)$ is an indicator function taking the values of 1 if the wind direction belongs

to $\mathcal{D}_t$ or 0 otherwise. Again, $f(x)$ is the common production frontier, so there is no differentiating subscript used on it. Let us take the difference between the two equations in (IV.2). Then, we have

$$\tilde{y}_{1-2}(x) = \tilde{\eta}_{2-1}(x) - \omega_1(x) \cdot \mathbb{1}_{\mathcal{D}_1}(x) + \omega_2(x) \cdot \mathbb{1}_{\mathcal{D}_2}(x) + \tilde{\epsilon}, \tag{IV.3}$$

where the tilde indicates a turbine difference term and the subscripts $1-2$ and $2-1$ signify the specific order of the difference. The above model is interpreted in a way that the power difference of Turbine 1 over Turbine 2 is due to the inherent production difference between the two turbines, $\tilde{\eta}_{2-1}(\cdot)$, and the power loss caused by the wake effect, characterized by either $\omega_1(\cdot)$ or $\omega_2(\cdot)$, both depending on specific wind conditions. Considering the mutual exclusiveness of $\mathcal{D}_1$ and $\mathcal{D}_2$, $\omega_1(\cdot)$ and $\omega_2(\cdot)$ will not appear at the same time.

To specify the above model, we further clarify what should be included in the input vector x. In general, it is well-known that the dominating input factors for wind power production are wind speed, V, and wind direction, D. However, Lee et al. (2015b) have shown that environmental factors other than wind speed and direction, such as air density and humidity, may also have a significant impact on wind power output. One advantage

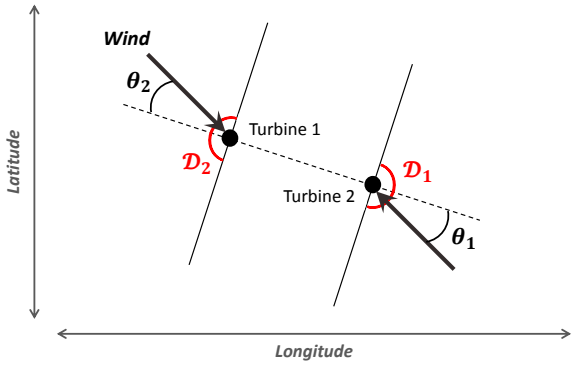

Figure IV.5: Two subsets of wind direction, $\mathcal{D}_1$ and $\mathcal{D}_2$. The union of the two subsets covers the entire 360° wind direction.

of using the power difference model (IV.3) is that we no longer need to consider other environmental factors because once we take the power difference between the two turbines, the impact of the environmental factors other than that of the wind is neutralized.

Furthermore, for two terms in equation (IV.3), either wind speed or wind direction, but not both, is needed as an input. The input to the indicator function $\mathbb{1}_{\mathcal{D}_1}(\cdot)$ or $\mathbb{1}_{\mathcal{D}_2}(\cdot)$ is apparently wind direction only. For the term of $\tilde{\eta}_{2-1}(\cdot)$, since it represents the between-turbine production difference independent of wake, we believe that it is only a function of wind speed, not a function of wind direction; the portion related to wind direction affecting the power difference $\tilde{y}_{1-2}$ or $\tilde{y}_{2-1}$ should be included in the wake related term ω.

As such, the power difference model can then be expressed as:

$$\tilde{y}_{1-2}(V, D) = \tilde{\eta}_{2-1}(V) - \omega_1(V, D) \cdot \mathbb{1}_{\mathcal{D}_1}(D) + \omega_2(V, D) \cdot \mathbb{1}_{\mathcal{D}_2}(D) + \tilde{\epsilon}. \tag{IV.4}$$

Here, $\tilde{\epsilon}$ is still an i.i.d. noise. We further assume it follows a normal distribution.

IV.3.3 Spline model with non-negativity and model estimation

In order to empirically estimate the power difference model in (IV.4), we need to assume an underlying model structure for the three functional terms, $\tilde{\eta}$, ω_1, and ω_2. For $\tilde{\eta}$ that has a single input, we choose to use a cubic smoothing spline function, whereas for the two wake power loss terms, ω_1 and ω_2, that have two inputs, we choose to use a thin plate spline model (Duchon, 1977), the multiple-dimensional generalization of the smoothing splines. As such, the power difference model (IV.4) becomes a spline model with the non-negativity constraint imposed on ω_1 and ω_2. We refer to the resulting model as the thin plate regression spline model with non-negativity (TPRS-N). Note that although the inefficiency term $\eta(\cdot)$ in the baseline power production model needs to be non-negative, $\tilde{\eta}(\cdot)$, which is the difference between two individual $\eta(\cdot)$'s, can take any value.

We understand that there are other options than the spline model structure for fitting

the power difference model in (IV.4), such as Gaussian process regression (Rasmussen and Williams, 2006) or kernel regression (Nadaraya, 1964; Watson, 1964). Without the non-negativity constraints, all these models produce outcome comparable with one another, so that which modeling option one chooses matters less. With the non-negativity constraints, we find it easier to work with the spline models.

To estimate the spline-based power difference model, we follow the generalized additive model (GAM) scheme (Hastie and Tibshirani, 1990). GAMs represent a univariate response as an additive sum of multiple smooth functions each having its own predictor variables. Estimation of GAMs can be performed by implementing the backfitting algorithm for which each smooth function is fitted for the residuals of all the others, iteratively one at a time until the fitted functions converge.

Consider n data pairs for which a residual r is paired with covariates x, i.e., (x_i, r_i) for $i = 1, \ldots, n$. Then, we estimate a smooth function by finding h that minimizes

$$\sum_{i=1}^{n} \{r_i - h(x_i)\}^2 + \lambda J[h], \tag{IV.5}$$

where λ is a penalty parameter controlling the trade-off between data fitting and smoothness of h. Both smoothing splines and thin plate splines solve the penalized least squares problem stated in (IV.5), but the measure of the smoothness, denoted by $J[h]$, would be different as they consider different dimensional functional spaces. For smoothing splines with a univariate predictor,

$$J[h] = \int_{\Re} h''(x)^2 dx,$$

whereas for thin plate splines with two predictors of x_1 and x_2,

$$J[h] = \int\int_{\Re^2} \left[\left(\frac{\partial^2 h(\boldsymbol{x})}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 h(\boldsymbol{x})}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 h(\boldsymbol{x})}{\partial x_2^2} \right)^2 \right] dx_1 \, dx_2.$$

The minimizer of (IV.5) for a cubic smoothing spline corresponds to a natural cubic spline with $n-2$ interior knots. If we use the B-spline basis to represent it, i.e., $h(x) = \sum_{j=1}^{n+2} \gamma_j B_j(x)$, equation (IV.5) can be rewritten in a matrix format as

$$(\boldsymbol{r} - \mathbf{B}\boldsymbol{\gamma})^T(\boldsymbol{r} - \mathbf{B}\boldsymbol{\gamma}) + \lambda \boldsymbol{\gamma}^T \boldsymbol{\Omega} \boldsymbol{\gamma}, \tag{IV.6}$$

where $\boldsymbol{r} = (r_1, \ldots, r_n)^T$, $\mathbf{B}_{ij} = B_j(x_i)$ and $\boldsymbol{\Omega}_{jl} = \int B_j''(x) B_l''(x) dx$, and $\boldsymbol{\gamma}$ is the coefficient vector of the $n+2$ basis functions, to be estimated. After taking the derivative of (IV.6) with respect to $\boldsymbol{\gamma}$ and setting the derivative equal to zero, we have

$$(\mathbf{B}^T\mathbf{B} + \lambda\boldsymbol{\Omega})\hat{\boldsymbol{\gamma}} = \mathbf{B}^T\boldsymbol{r}.$$

Let $\mathbf{M} = (\mathbf{B}^T\mathbf{B} + \lambda\boldsymbol{\Omega})$, and calculate its Cholesky decomposition $\mathbf{M} = \mathbf{L}\mathbf{L}^T$. Then, solving $\mathbf{L}\mathbf{L}^T\hat{\boldsymbol{\gamma}} = \mathbf{B}^T\boldsymbol{r}$ by back-substitution provides $\hat{\boldsymbol{\gamma}}$ and thereby $\hat{h}(x) = \sum_{j=1}^{n+2} \hat{\gamma}_j B_j(x)$ in $O(n)$ operations.

When using thin plate splines, the solution of (IV.5) is equivalent to that of

$$\min ||\boldsymbol{r} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Phi}\boldsymbol{\delta}||^2 + \lambda \boldsymbol{\delta}^T \boldsymbol{\Phi} \boldsymbol{\delta}, \quad \text{subject to } \mathbf{X}^T\boldsymbol{\delta} = \mathbf{0}. \tag{IV.7}$$

where the $n \times 3$ matrix $\mathbf{X} = [\mathbf{1}_n; \boldsymbol{x}_1; \boldsymbol{x}_2]$ includes the unit vector of size n as its first column and the n observations for the two covariates as its second and third columns. The radial basis matrix $\boldsymbol{\Phi}$ is defined by $\boldsymbol{\Phi}_{ji} = \phi(||x_j - x_i||) = ||x_j - x_i||^2 \log||x_j - x_i||$ for $i, j = 1, \ldots, n$. The 3-dimensional vector $\boldsymbol{\beta}$ and the n-dimensional vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$ are, respectively, the coefficients associated with $\mathbf{X}$ and those associated with the radial basis functions, both sets of coefficients to be estimated.

Using as many basis functions as the number of data points could be computationally challenging when the data size n is considerably large. Different from the univariate spline

problem that can be solved by $O(n)$ operations as discussed above, the computations for the thin plate splines require $O(n^3)$ operations (Hastie and Tibshirani, 1990). To overcome the computational problem, Wood (2003) proposed the thin plate regression splines (TPRS). Although the name includes the term 'regression splines,' unlike other regression splines, TPRS does not require the selection of the knots. To improve the computational efficiency, TPRS uses only k eigenbasis ($k \ll n$) corresponding to the largest k eigenvalues of the basis matrix $\mathbf{\Phi}$, and consequently reduces the rank of the basis matrix significantly.

TPRS can be fitted as follows. First, by applying the eigen decomposition of $\mathbf{\Phi}$, we have $\mathbf{\Phi} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{\Phi}$ and arranged in a non-increasing order, i.e., $\mathbf{D}_{i,i} \geq \mathbf{D}_{i+1,i+1}$ for $i = 1, \ldots, n-1$. $\mathbf{U}$ is an orthogonal matrix whose columns are the eigenvectors ordered accordingly. Then, TPRS considers the first k columns of the matrix $\mathbf{U}$, denoted by $\mathbf{U}_k$, resulting in a rank k eigenbasis matrix $\mathbf{\Phi}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{U}_k^T$ where $\mathbf{D}_k$ is a $k \times k$ diagonal matrix taking the first k rows and columns of $\mathbf{D}$. Subsequently, the constraint in (IV.7) can be dropped by using QR decomposition on $\mathbf{U}_k^T\mathbf{X}$, i.e., $\mathbf{U}_k^T\mathbf{X} = \mathbf{Q}\mathbf{R}$ where $\mathbf{Q}$ is a $k \times k$ orthogonal matrix and $\mathbf{R}$ is a $k \times 3$ upper triangular matrix. Let $\mathbf{Z}_k$ take the last $k-3$ columns of $\mathbf{Q}$. By restricting $\boldsymbol{\delta} = \mathbf{U}_k\mathbf{Z}_k\boldsymbol{\delta}_k$ with $(k-3)$-dimensional coefficient vector $\boldsymbol{\delta}_k$, the rank k approximation then can be used to fit TPRS by solving

$$\min ||\boldsymbol{r} - \mathbf{X}\boldsymbol{\beta} - \mathbf{U}_k\mathbf{D}_k\mathbf{Z}_k\boldsymbol{\delta}_k||^2 + \lambda\boldsymbol{\delta}_k^T\mathbf{Z}_k^T\mathbf{D}_k\mathbf{Z}_k\boldsymbol{\delta}_k, \tag{IV.8}$$

for the unknown $\boldsymbol{\beta}$ and $\boldsymbol{\delta}_k$. Then, prediction for any given x can be achieved by calculating $\hat{\boldsymbol{\delta}} = \mathbf{U}_k\mathbf{Z}_k\hat{\boldsymbol{\delta}}_k$ and plugging $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\beta}}$ into

$$\hat{h}(x) = \mathbf{X}\hat{\boldsymbol{\beta}} + \sum_{i=1}^{n} \hat{\delta}_i\phi(||x - x_i||). \tag{IV.9}$$

Recall that we assume the wake power loss term ω_t is non-negative so that our model can be consistent with physical understanding, but the modeling procedure of TPRS does not guarantee the non-negativity. In order to make sure the wake power loss is indeed non-negative, we decide to apply an exponential transformation on top of the conventional TPRS estimation in (IV.9), such as

$$\hat{\omega}(x) = \exp\{\mathbf{X}\hat{\boldsymbol{\beta}} + \sum_{i=1}^{n} \hat{\delta}_i \phi(||x - x_i||)\}. \tag{IV.10}$$

Because of this change, instead of (IV.8), we solve

$$\min ||\boldsymbol{r} - \exp\{\mathbf{X}\boldsymbol{\beta} + \mathbf{U}_k\mathbf{D}_k\mathbf{Z}_k\boldsymbol{\delta}_k\}||^2 + \lambda\boldsymbol{\delta}_k^T\mathbf{Z}_k^T\mathbf{D}_k\mathbf{Z}_k\boldsymbol{\delta}_k, \tag{IV.11}$$

with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\delta}_k$.

One may argue that once the function form is expressed in the exponential form in (IV.10), the penalty term in (IV.11) needs to be re-derived, following the formulation of (IV.5) and through calculating the second derivatives of the new exponential term in (IV.10). Doing so would make the regularized learning formulation in (IV.11) messier but the benefit is marginal. Recall the role of the penalty term is to impose certain degree of smoothness in the final estimation. By using the same penalty term as in (IV.8), we understand that the balance between the goodness-of-fit and the smoothness may not be optimal. Regardless of that, however, the optimality would not be attained anyway since we use the basis truncation to speed up the computation (i.e., the use of k eigenvalues). The inclusion of the penalty term still ensures smoothness in the final fit, and through the exponentiation, the final model is smoother. Due to these considerations, the treatment we apply here was in fact advocated previously in smoothing spline research (Ramsay and Silverman, 2005) as an effective way to handle non-negativity.

When estimating a GAM, one generally considers a constant term as the global mean. In fact, the global mean is typically calculated and subtracted in advance, before implementing the backfitting algorithm that estimates the rest of the functional terms. In the power difference model (IV.4), this constant term should be part of the turbine-difference term $\tilde{\eta}(\cdot)$, meaning that a portion of the turbine difference is constant regardless of the wind conditions, while the other portion may change with the wind speed. For the implementation of the GAM backfitting algorithm, we can re-write equation (IV.4) as

$$\tilde{y} = \alpha + [\tilde{\eta}(V) - \alpha] - \omega_1(V, D) \cdot \mathbb{1}_{\mathcal{D}_1}(D) + \omega_2(V, D) \cdot \mathbb{1}_{\mathcal{D}_2}(D) + \tilde{\epsilon}, \tag{IV.12}$$

and estimate α using the global mean and $[\tilde{\eta}(V) - \alpha]$ using a cubic smoothing spline (and the wake loss terms using TPRS-N). Afterwards, $\tilde{\eta}(V)$ is restored by $\hat{\alpha} + \hat{\eta}(V)$.

Before implementing the backfitting algorithm, some tuning parameters need to be set, including the smoothing parameter λ and the value of the reduced rank k used for improving computational efficiency of TPRS-N. There are in fact three smoothing parameters λ, one for each smooth function estimation, associated with $\tilde{\eta}(\cdot)$ and the two $\omega(\cdot)$'s, respectively. They are chosen based on a 10-fold cross validation while applying grid search. For the choice of k, Wood (2003) stated that its choice is not so critical as long as it is larger than the degrees of freedom required for the estimation. In the subsequent analysis sections, we set $k = 30$ which, we believe, is large enough for the wake effect analysis application based on our graphical inspection of the estimation results. Finally, we set a threshold τ determining the convergence of the model fitting to 0.1 which is a sufficiently small number considering that the magnitude of the functional estimates changes exponentially due to the imposition of non-negativity.

The backfitting algorithm for the power difference model is summarized in Algorithm 2.

Algorithm 2 Backfitting algorithm for wake power loss estimation.

Initialize:

$m \leftarrow 0$; $\hat{\alpha} \leftarrow \sum_{i=1}^{n} y_i/n$; $\hat{\boldsymbol{\eta}}^m \leftarrow \mathbf{0}$; $\hat{\boldsymbol{\omega}}_1^m \leftarrow \mathbf{0}$; $\hat{\boldsymbol{\omega}}_2^m \leftarrow \mathbf{0}$.

repeat

Set $m \leftarrow m + 1$.

Estimation of $\hat{\boldsymbol{\eta}}$

Calculate partial residuals: $\boldsymbol{r}_\eta \leftarrow \boldsymbol{y} - \hat{\alpha} + \hat{\boldsymbol{\omega}}_1^{m-1} - \hat{\boldsymbol{\omega}}_2^{m-1}$.

Set $\hat{\boldsymbol{\eta}}^m$ by fitting smoothing spline to $\boldsymbol{r}_\eta$ with respect to $\boldsymbol{V}$.

Estimation of $\hat{\boldsymbol{\omega}}_1$

Calculate partial residuals: $\boldsymbol{r}_{\omega_1} \leftarrow -(\boldsymbol{y} - \hat{\alpha} - \hat{\boldsymbol{\eta}}^m - \hat{\boldsymbol{\omega}}_2^{m-1})$.

Set $\hat{\boldsymbol{\omega}}_1^m$ by fitting thin plate regression spline with non-negativity to $\boldsymbol{r}_{\omega_1}$ with respect to $\boldsymbol{V}$ and $\boldsymbol{D}$ for the data whose $D \in \mathcal{D}_1$.

Estimation of $\hat{\boldsymbol{\omega}}_2$

Calculate partial residuals: $\boldsymbol{r}_{\omega_2} \leftarrow \boldsymbol{y} - \hat{\alpha} - \hat{\boldsymbol{\eta}}^m + \hat{\boldsymbol{\omega}}_1^m$.

Set $\hat{\boldsymbol{\omega}}_2^m$ by fitting thin plate regression spline with non-negativity to $\boldsymbol{r}_{\omega_2}$ with respect to $\boldsymbol{V}$ and $\boldsymbol{D}$ for the data whose $D \in \mathcal{D}_2$.

Computation of convergence criterion

$$\Delta \leftarrow \frac{||\hat{\boldsymbol{\eta}}^m - \hat{\boldsymbol{\eta}}^{m-1}|| + ||\hat{\boldsymbol{\omega}}_1^m - \hat{\boldsymbol{\omega}}_1^{m-1}|| + ||\hat{\boldsymbol{\omega}}_2^m - \hat{\boldsymbol{\omega}}_2^{m-1}||}{||\hat{\boldsymbol{\eta}}^{m-1}|| + ||\hat{\boldsymbol{\omega}}_1^{m-1}|| + ||\hat{\boldsymbol{\omega}}_2^{m-1}||}.$$

until $\Delta \leq \tau$ where τ is a pre-specified threshold.

IV.4 Performance comparison of different wake models

In this section, we compare the proposed spline-based wake model with other wake models in terms of the prediction error of the power difference. Because it is extremely

difficult, if not impossible, to measure the actual wake power loss directly, prediction or estimation of the power difference becomes an important proxy alluding to model capability accounting for wake effects in wind power production. The same proxy was used by You et al. (2017) for evaluating their wake model. Furthermore, power difference prediction is in and by itself useful in a number of wind energy applications, for instance, to quantify the effect of a wind turbine upgrade action through the comparison of a pair of turbines (Lee et al., 2015a).

For this model comparison, we use actual operational data collected from an onshore wind farm in the US. The wind farm houses more than 200 wind turbines and four meteorological mast towers spreading over a relatively large area. From this wind farm, we take six pairs of wind turbines (in total, 12 wind turbines) into consideration. The turbine pairs are chosen such that no other turbines except them are located within 10 times the rotor diameter d. Such arrangement is to find a pair of turbines that are free of other turbine's wake, so that the wake analysis result can be reasonably attributed to the other turbine in its own pair. Ammara et al. (2002) stated that theoretical and experimental studies have generally suggested that wake velocity deficit is minimal beyond $10d$ downstream of a wind turbine. As such, we expect the wake power loss due to other turbines becomes virtually negligible by the $10d$ restriction.

Figure IV.6 shows the relative locations of the six pairs of turbines on the wind farm. The circle around each turbine indicates the $10d$ radius from the turbine. All turbine pairs happen to have the northwestern-to-southeastern orientation. So we designate the turbine on the northwestern side as Turbine 1 and the one on the southeastern side as Turbine 2 for all turbine pairs. Table IV.1 provides the between-turbine distances, in terms of a multiple of the rotor diameter, and the relative positional angles between a pair of turbines. Following the wind direction notation illustrated in Figure IV.5 and based on the specific relative positions between a pair of turbines, we can divide wind direction into two distinct

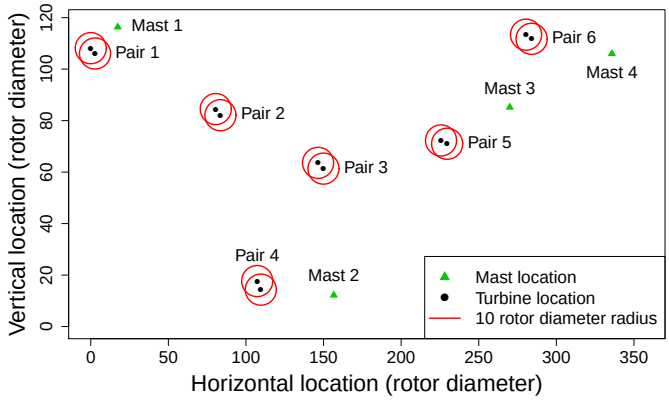

Figure IV.6: Locations of the six pairs of wind turbines and the four meteorological masts. The distances along both axes are expressed as a multiple of the rotor diameter of the turbines. All turbines have the same rotor diameter.

sectors of $\mathcal{D}_1$ and $\mathcal{D}_2$ for each turbine pair. For wind direction $D \in \mathcal{D}_2$, Turbine 1 is wake free and Turbine 2 is in the wake, whereas for $D \in \mathcal{D}_1$, Turbine 2 is wake free and Turbine 1 is in the wake.

We have a year worth of the operational data of the six pairs of turbines, taken during roughly a year-long period between 2010 and 2011. The datasets include wind power output, wind speed, and wind direction. The data provided to us are recorded for every 10 minutes, and they are the averages of the raw measurements calculated over distinct 10-minute time intervals, following the IEC standard (IEC12.1, 2005). Overall, it gives us approximately 33,500–38,500 data pairs for one year after accounting for the missing val-

Table IV.1: Between-turbine distances and relative positions of the six pairs of turbines. Bearing 1 to 2 indicates a relative direction of Turbine 1 to the location of Turbine 2, and Bearing 2 to 1 is similarly defined.

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
Between-turbine distance	$3.4d$	$3.8d$	$4.2d$	$4.1d$	$4.1d$	$3.9d$
Bearing 1 to 2 (°)	307.1	308.7	302.6	325.0	288.3	294.2
Bearing 2 to 1 (°)	127.1	128.7	122.6	145.0	108.3	114.2

ues and eliminating data records outside the normal operating range (say, with a negative power output).

The wind power outputs are measured on a turbine. The wind measurements, including wind speed and direction, are measured at the nearby mast tower. Wind speed is also measured on the turbines, but not the wind direction. Considering that there are only four mast towers on this wind farm and some of the turbine pairs are relatively far away from any of the mast towers, we decide to use the wind speed measurements obtained at the turbines. Wind direction is only available at the mast towers, so the mast measurements are used as there is no other option. Of course, we take the data from the mast closest to a turbine pair. Specifically, we use the data from Mast 1 for Pairs 1 and 2, Mast 2 for Pairs 3 and 4, and Mast 3 for Pair 5 and 6.

Note that the wind speed measurements on a turbine is obtained after the rotor, namely in its own wake. IEC 61400-12-2 (IEC12.2, 2013) recommends dealing with this issue through a nacelle transfer function (NTF), which describes the relation between the free inflow wind speed and that measured at the turbine anemometer in the wake. The wind operator who provided us the data informs us that the turbine wind speed measurements have been adjusted through a NTF and can be treated as if they were measured in front of the rotor.

We evaluate the performance of a model with respect to its out-of-sample prediction errors. For this, we split each turbine pair's annual data into training and testing subsets by a ratio of 80:20. In other words, we use randomly selected 80% of a given dataset to train the model and the rest 20% to calculate the prediction error. To measure the prediction error, we use the root mean squared error (RMSE) defined by

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} (\tilde{y}_i - \hat{\tilde{y}}_i)^2},$$

where $\tilde{y}_i$ for $i = 1, \ldots, n_{\text{test}}$ is the power difference calculated for the testing dataset, $\hat{\tilde{y}}_i$ is the corresponding prediction of the power difference, and n_{test} is the number of data pairs in a test set, typically between 6,000 and 8,000 in amount.

We compare the proposed method with three other methods: a physics-based method and two data-driven models. The physics based method is Jensen's model (Jensen, 1983). The two data-driven models are the data binning approach (Barthelmie et al., 2010) and the GMRF wake model (You et al., 2017). We choose the Jensen's model as the representation of the physics-based model because it is widely used in the commercial operation environments. The CFD-based models are still limited in restrictive research settings that have access to supercomputing capability and thus not practically applicable to commercial sized wind turbines in actual wind farm operations.

To implement these methods, certain tuning parameters are set. The Jensen's model has a wake decay constant. We set this value to 0.075, the value commonly used for an onshore wind farm (WAsP, 2016). The implementation of the physics-based model additionally requires power curve estimation for which we follow the standard procedure described in IEC12.1 (2005). For the data binning approach, instead of using a subset of data corresponding to a restricted range of wind speed and wind direction, we simply choose to use the entire dataset. This extension is straightforward, and the result of using the whole dataset is better in terms of RMSE than using the restrictive subset. We generate wind direction bins using a bin width resolution of 5° following Barthelmie et al. (2010). For the GMRF model, the authors of You et al. (2017) have generously implemented their method on our data (all six turbine sets) and reported the resulting RMSE values to us.

Table IV.2 presents the RMSE values for the four methods and six turbine pairs, denoted as percentage values of the maximum power output. The proposed spline-based model is labeled as 'GAM with TPRS-N' in the table. The Jensen's model, due to its simplicity and additional errors induced by the use of a power curve based conversion, leaves

Table IV.2: Comparison of prediction error in terms of RMSE. The value in the table is the percentage of power difference relative to the maximum power of the turbine. The boldface values are the smallest in each column.

	RMSE					
	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
Jensen's model	11.03%	8.87%	11.09%	9.71%	9.56%	10.20%
GMRF	8.46%	7.52%	8.88%	7.97%	7.98%	8.77%
Data binning approach	7.78%	6.67%	8.18%	8.00%	7.06%	7.51%
GAM with TPRS-N	**6.72%**	**6.27%**	**8.00%**	**7.59%**	**6.82%**	**7.06%**

a relatively large portion of variation in the original data unexplained and hence registers the highest RMSE values. Relative to the Jensen's model, all data-driven methods significantly reduce the level of uncertainty by accounting for the variation observed in the data.

Recall that the GMRF model was not specifically developed for the single wake situations. By construction, the GMRF model is designed to perform better with more turbines since it benefits from the spatial modeling of multiple turbines at different locations. Understandably, the method loses its benefits when applied to a pair of turbines. Still, the method shows significant improvement with an 18% reduction in prediction error on average compared to the Jensen's model.

The data binning approach, to fit the trend of data without any restriction, in fact attains competitive prediction errors. This should not come as a surprise, as the binning approach is an extreme version of a localized fitting method and can adapt to local data features, as long as one uses a small enough binning resolution and there are dense enough data points to fit. The data binning approach is often competitive or the best method in terms of out-of-sample RMSE value when comparing with other fitting methods. The fact that its RMSE is in fact larger than that of our proposed spline-based model suggests that the data binning approach overfits the (training) data. We also reiterate that the binning approach

is less insightful at providing wake characteristic, as we argued in an early section. In addition, the data binning approach will run into a dimensionality issue rather quickly, if there are more variables to be binned than the wind direction, an issue that has been discussed at length in Lee et al. (2015a).

Our proposed model demonstrates its superiority over other alternatives in terms of the prediction error of the power difference. It yields the smallest RMSE values across all six turbine pairs, and its RMSE is, on average, 30% smaller than that of the Jensen's model, 14% smaller than the GMRF model, and 6% smaller than the data binning approach.

IV.5 Analysis of wind turbine wakes in actual operations of wind turbines

In this section, we quantify annual wake power loss in actual wind turbine operations. Quantification of the wake power loss based on an annual period supports economic assessment of wake effect in terms of annual energy production (AEP), a key performance metric that is contractual binding and recommended by IEC (IEC12.1, 2005). Doing so also provides practical insights into the economic impact of decisions and actions attempting to alleviate the wake power loss.

To quantify the annual wake power loss and derive the wake characteristics revealed during an annual period, we apply the proposed method to the entire year-long dataset. Figure IV.7 illustrates the fitted wake effect. By our estimation, the wake loss is strictly positive, but what we show in the plot is actually $-\hat{\omega}_1(V, D) \cdot \mathbb{1}_{\mathcal{D}_1}(D) + \hat{\omega}_2(V, D) \cdot \mathbb{1}_{\mathcal{D}_2}(D)$, so that one sees both positive and negative portions. The power difference of some pairs of turbines, when plotted against wind direction, exhibits large variation, with several peaks and troughs. Even under such a noisy circumstance, our model captures the wake power loss signals well by focusing on where the wake power loss is expected. In the figure, the dashed lines indicate the bearings, i.e., $\theta_1 = 0$ and $\theta_2 = 0$. Comparing Figure IV.7(e) to Figure IV.4 (generated from Pair 5), it is obvious that our wake loss estimation method

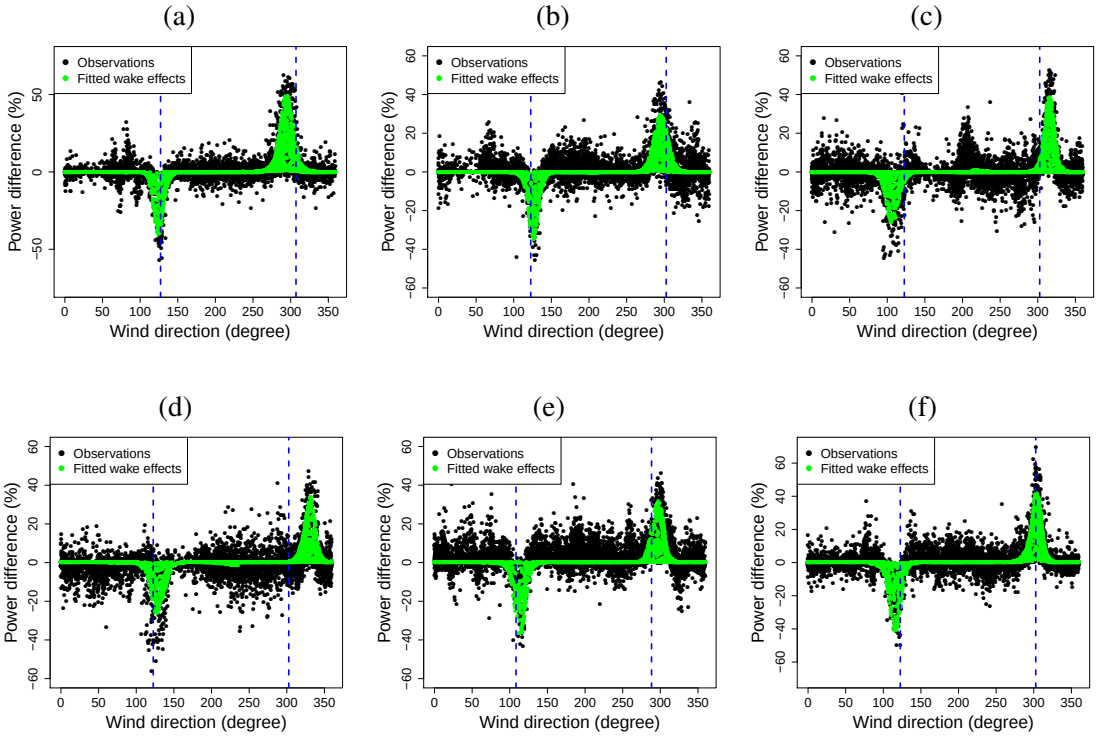

Figure IV.7: Estimated wake effects using GAM with TPRS-N: (a) for Pair 1; (b) for Pair 2; (c) for Pair 3; (d) for Pair 4; (e) for Pair 5; (f) for Pair 6. The green areas represent the fitted wake loss in terms of $-\hat{\omega}_1(V,D)\cdot\mathbb{1}_{\mathcal{D}_1}(D)+\hat{\omega}_2(V,D)\cdot\mathbb{1}_{\mathcal{D}_2}(D)$. Two dashed vertical lines indicate wind direction that is parallel to the line connecting the pair of turbines.

captures the signals much better than the data binning approach could, making the subsequent derivation of the wake characteristics easier. We also observe from Figure IV.7 that the wind direction associated with the highest power loss is not exactly aligned with the bearings of the turbine pairs. This implies that there is measurement error in wind direction. In the past, when practitioners apply the data binning strategy, a common approach is to generate angle bins starting around the bearings (a bearing would be the midpoint of the angle bin determining the wake depth) and then propagate with a resolution of 5 degrees. It turns out that such a practice has an obvious disadvantage and will surely underestimate the wake depth, as the data points in the bin including a bearing could have lower power

loss than those around the actual peak angle and this would considerably decrease the sample average that goes into the estimate.

Table IV.3 shows the wake characteristics for the six turbine pairs. The first two rows are the wake depths, namely the magnitude of the wake power losses. The last two rows are the wake widths. The wake depth is identified as the peak corresponding to the maximum power loss. The wake width is supposed to be determined by the angles around the bearings at which the power loss eventually becomes zero. However, given noisy signals spreading over a large range of wind directions, the fitted wake power loss is not completely zero even in the regions where it is unquestionably wake free. To estimate the wake width, we therefore find the range of wind direction around the bearings whose loss is greater than 1% of the rated power of the turbine.

For the wake depth, Table IV.3 presents two percentage values for each turbine. The one outside the parenthesis is the wake power loss relative to the rated power of that turbine whereas the one inside the parenthesis is the loss relative to the free stream equivalent power output. Recall that throughout the chapter, we do not show the actual power values due to the confidentiality agreement in place but show the normalized power values, in a percentage, relative to the rated power. These turbines belong to the general 2 MW turbine class. It does not mean that the rated power of the turbines is exactly 2 MW but it is in that vicinity. Using this information, one can estimate the wake power loss magnitude in the unit of mega-watts.

In the literature, however, the wake power loss is often expressed as the ratio of the

Table IV.3: Wake depth and width for the six pairs of turbines.

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
Depth: Turbine 1	39.7% (62.7%)	34.5% (41.6%)	26.3% (35.9%)	26.2% (39.1%)	37.9% (49.3%)	41.0% (56.6%)
Depth: Turbine 2	49.0% (57.7%)	29.1% (36.0%)	38.5% (47.9%)	33.3% (47.4%)	31.6% (57.2%)	41.4% (57.3%)
Width: Turbine 1	39.2°	44.0°	47.4°	52.3°	45.1°	48.8°
Width: Turbine 2	61.1°	49.7°	40.0°	46.6°	43.0°	48.9°

loss over the free stream equivalent power output (Barthelmie and Jensen, 2010; Hansen et al., 2012; Adaramola and Krogstad, 2011), so we have

$$\frac{\hat{\omega}_t(V_i, D_i)}{\hat{y}_t(V_i, D_i) + \hat{\omega}_t(V_i, D_i)}, \quad t = 1, 2, \quad i = 1, \ldots, n,$$

where $\hat{y}_t(V_i, D_i)$ denotes the expected power generation given (V_i, D_i). Depending on (V_i, D_i), $\hat{y}_t(V_i, D_i)$ could be the expected power in the wake of another turbine, so that we mean to recover the free stream equivalent power output by adding $\hat{y}_t(V_i, D_i)$ and $\hat{\omega}_t(V_i, D_i)$. To calculate $\hat{y}_t(V_i, D_i)$, we first define a neighborhood of (V_i, D_i), i.e., $\mathcal{N}_i = \{(V, D) : V \in (V_i - \epsilon_V, V_i + \epsilon_V], D \in (D_i - \epsilon_D, D_i + \epsilon_D]\}$ where ϵ_V and ϵ_D are predetermined constants. We set $\epsilon_V = 0.25$ m/s and $\epsilon_D = 2.5°$, following the common practice. Then, we calculate $\hat{y}_t(V_i, D_i)$ by taking the average of the power outputs whose corresponding wind speed and direction is a member of $\mathcal{N}_i$. This is a two-dimensional binning with $2\epsilon_V$ and $2\epsilon_D$ as the respective bin width. The second percentage values in Table IV.3, namely the ones inside the parenthesis, are the wake power loss expressed in this conventional fashion.

The peak power loss relative to the free stream equivalent (the value inside the parenthesis) ranges from 36% to 63%. The wake width for the 12 turbines ranges from $39°$ to $61°$ with concentration around $40°$–$50°$. The commonly stated wake depth in the literature is in the range of 30%–40% (Barthelmie et al., 2010; Bhutta et al., 2012; Sanderse et al., 2011), which appears to be at the lower bound of our estimates. In addition, our wake width estimates are also noticeably larger than the $25°$ to $40°$ range stated previously (Barthelmie et al., 2010; McKay et al., 2013; Troldborg et al., 2011). We believe that the difference can be attributed to two major factors. The first one is that our estimation can identify the wake region more accurately, producing better estimates of the two main characteristics, whereas the methods in the literature rely on *ad hoc* data segmentation and

partition and often use a partial set of data based on a pre-selected wind direction, and consequently, their wake power loss estimate does not capture the characteristics as well as our method does. The second factor is that the historical estimates are usually the average over multiple turbines, understandably leading to a range narrower.

Table IV.4 shows how each term in the power difference model (IV.4) affects the power generation of a turbine pair in an annual period, namely the AEP power difference or AEP loss. The first row is the between-turbine power production difference independent of wake effect, expressed relative to the rated power. The second and third rows present the wake loss. Similarly to Table IV.3, the values outside the parentheses is the loss relative to the rated power, whereas the values inside the parentheses is the loss relative to the free stream equivalent. Both percentages represent the AEP wake loss but use different baselines.

The wake loss relative to the rated power is in fact related to the capacity factor of a wind turbine (Wikipedia, 2017). Recall that the capacity factor is the ratio of the actual power production of a turbine for a selected period of time, say, one year, over the supposed power production if the turbine always operates at its maximum capacity (i.e., at the rated power) during the same period. The typical range of the capacity factor is 25%–30%. The wake loss relative to the rated power can be seen as the direct reduction to a turbine's capacity factor. We hereby refer to the corresponding AEP loss as the capacity factor AEP loss, and refer to the AEP loss relative to the free stream equivalent as the traditional AEP

Table IV.4: Annual power loss for the six turbine pairs.

	Percentage measure (%)					
	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6
Turbine difference ($\tilde{\eta}_{2-1}$)	0.22	1.04	0.22	-1.45	3.38	0.85
Wake loss: Turbine 1 (ω_1)	0.83 (1.65)	0.73 (1.55)	0.50 (1.12)	0.61 (1.37)	0.63 (1.38)	0.68 (1.40)
Wake loss: Turbine 2 (ω_2)	2.00 (4.09)	1.24 (2.71)	1.30 (3.01)	1.10 (2.41)	0.98 (2.33)	1.80 (3.83)
Average loss for the pair	1.42 (2.86)	0.98 (2.12)	0.90 (2.06)	0.86 (1.89)	0.81 (1.84)	1.24 (2.59)

loss, which is computed, if using Turbine 1 group as an example, by

$$\frac{\sum_{i=1}^{n} \hat{\omega}_1(V_i, D_i)}{\sum_{i=1}^{n} \{\hat{y}_1(V_i, D_i) + \hat{\omega}_1(V_i, D_i)\}}.$$

The fourth row is the average AEP loss for a pair of turbines. The average is weighted by the number of data points in the respective wake regions to account for the annual distribution of the AEP loss for the turbine pairs. For this reason, the values in the fourth row may be slightly different from the simple average of the two individual losses. The traditional AEP loss for a pair is computed by

$$\frac{\sum_{i=1}^{n} \{\hat{\omega}_1(V_i, D_i) + \hat{\omega}_2(V_i, D_i)\}}{\sum_{i=1}^{n} \{\hat{y}_1(V_i, D_i) + \hat{\omega}_1(V_i, D_i) + \hat{y}_2(V_i, D_i) + \hat{\omega}_2(V_i, D_i)\}}.$$

The average capacity factor AEP loss is computed by setting the denominator in the above equation to be $\sum_{i=1}^{n} \{(\text{rated power}) + (\text{rated power})\} = 2n \cdot (\text{rated power})$.

From Table IV.4, one may notice that the magnitude of the between-turbine difference is sizeable, sometimes even larger than the wake effect. This result suggests that modeling of the between-turbine difference as a separate term in the power difference model is important to our mission to estimate the wake effect; otherwise, the estimate of the wake effect can be biased considerably.

One immediately observes that the AEP losses are much smaller than the peak power loss (wake depth). This is expected because the annual loss is the average over all kinds of wind speed and direction conditions in an entire year. Under many circumstances, the wake loss is much smaller than the peak loss. The capacity factor AEP loss is between 0.5–2.0%, meaning that if the turbine's actual capacity factor is 25%, then its ideal capacity factor, if the turbine always operates wake free, could have been 25.5% to 27%. This difference, while appearing as a small percentage, should not be taken lightly. Consider a

wind farm housing 200 turbines all in the 2 MW turbine class. A 1% capacity factor AEP loss for the whole farm translates to $1.3 million loss in revenue at the wholesale price of $37 per MWh (Statista, 2016).

One may also notice that the wake loss endured by Turbine 2 in a pair is always greater than that of Turbine 1. This can be explained by the relative position of the turbine and the prevailing wind direction over this farm during that particular year. Figure IV.8 presents the wind rose plots for three pairs of the turbines. The plots show that the north-western wind, for which Turbine 2 of each pair endures power loss, is more frequent and stronger than the south-eastern wind for which Turbine 1 experiences power loss. Unsurprisingly, we observe the AEP loss of Turbine 1 group is usually less than 0.83% (1.65%), whereas the AEP loss for Turbine 2 group is greater than 0.98% (2.33%) and can be as high as 2.00% (4.09%).

In the literature, it is well-known that turbine spacing is a decisive factor affecting the magnitude of wake power loss (Laan et al., 2015; Sanderse et al., 2011; Barthelmie and Jensen, 2010). We therefore suspect that the variation of the annual power loss among the

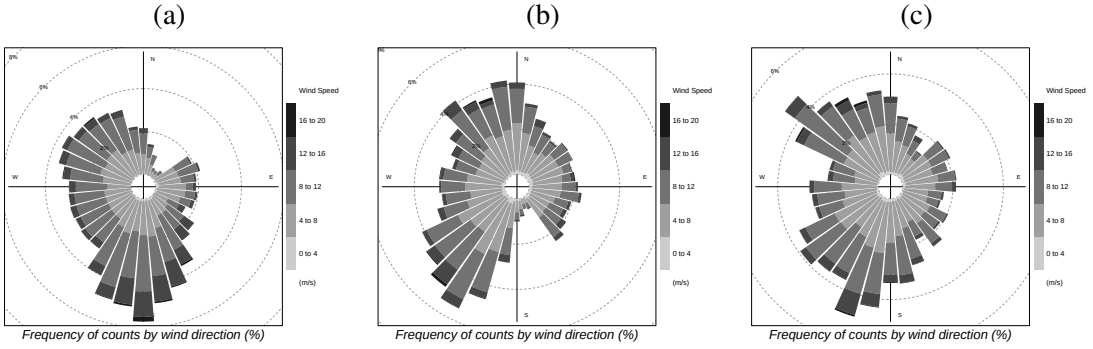

Figure IV.8: Wind rose plots illustrating the relative frequency of incoming wind for different direction sectors and for different speed ranges: (a) for Pair 1; (b) for Pair 3; (c) for Pair 5. These wind rose plots of Pair 1, 3, and 5 are representative of those of Pair 2, 4, and 6, respectively, because such group of two pairs shares the same met mast that takes the wind direction measurements.

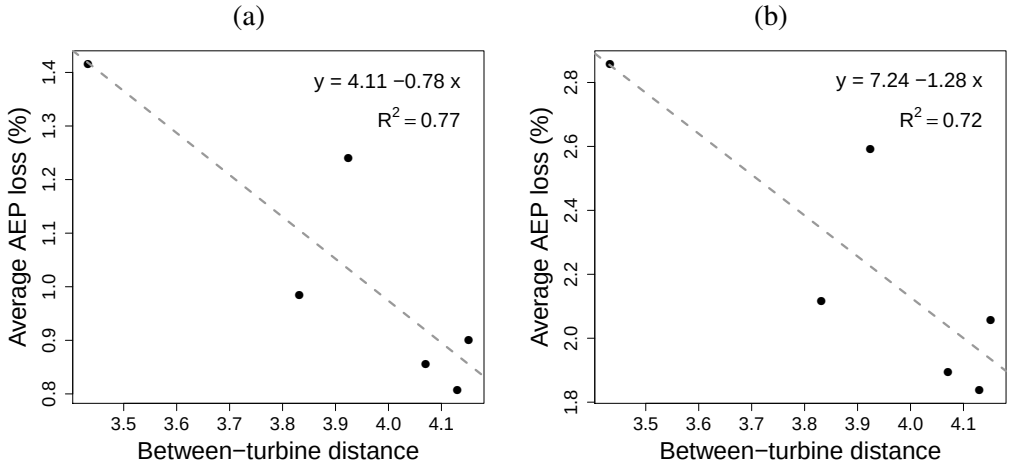

Figure IV.9: Relation between AEP losses and turbine spacing: (a) for the capacity factor AEP loss; (b) for the traditional AEP loss. The between turbine distance is expressed as a multiple of the rotor diameter.

individual turbine pairs can be explained by the between-turbine distance. Using the average AEP loss for the six turbine pairs (the fourth row in Table IV.4) and the corresponding between-turbine distances, we fit a simple linear regression model as has been done by Barthelmie and Jensen (2010). Figure IV.9 shows the scatter plots and the regression line fitting a respective AEP loss. For the capacity factor AEP loss, the p-values of the intercept and slope estimate are 0.008 and 0.022, respectively. For the traditional AEP loss, the corresponding p-values are 0.010 and 0.034. These results confirm that the turbine spacing indeed by and large explains the pair-wise difference in AEP losses. Additionally, an extrapolation based on the fitted regression lines suggests that the wake loss would diminish after the turbine spacing reaches either $5.2d$ or $5.7d$, depending on which AEP loss is used in the analysis. Nevertheless, in either circumstance, the $10d$ separation used in this study to isolate a particular turbine pair from the rest of turbines appears safe enough to render the turbine pairs free of wake of any other turbine on the wind farm.

Regressing the turbines' inherent production difference (the first row in Table IV.4)

on the between-turbine distance, on the other hand, suggests that there is no significant correlation between them. The p-values of the intercept and slope estimate in this case are 0.81 and 0.77, respectively, with R^2 of 0.02. As such, unlike the wake effect, the between-turbine production difference does not seem to be a function of the between-turbine distance. The obvious distinction of the regression results between the two cases supports that our model is capable of separating the power difference ($\tilde{y}$) into the between-turbine production difference part (i.e., $\tilde{\eta}$) and the wake effect part (i.e., ω_1 and ω_2), so that our wake effect estimate is well derived.

IV.6 Concluding remarks

Our study presents a data-driven wake effect model based on a spline model structure. A non-negativity constraint is incorporated in the model estimation to make sure that the estimate of wake power loss is consistent with the physical understanding. Our spline-based wake model produces the smallest prediction error when compared with one physics-based wake model and two data-driven wake models. More importantly, the proposed wake model produces the estimate of important wake characteristics, including wake depth, wake width, and annual wake power loss, which are useful to future designs and ongoing operations of wind turbines.

Application of the proposed wake model is limited to the analysis of single wake behaviors because of its model structure established upon pair-wise differencing between two turbines' power output. The general power production function introduced prior to the pair-wise analysis does allow a potential extension to multiple turbines case at the wind farm level. Another possibility is to combine the strength of this spline wake model with that of You et al. (2017) and develop a wake model that accounts for spatial relationship among multiple turbines while imposing non-negativity on wake loss at the same time. Such data-driven models could be promising alternatives to complement compli-

cated physics-based wake models for evaluating the wake power loss of a wind farm.

CHAPTER V

CONCLUSIONS

In this chapter, we summarize all the development achieved in this dissertation and highlight contributions in a broader perspective. We also discuss potential extensions of the research developed in this dissertation.

V.1 Summary

This dissertation has focused on improving the performance evaluation practice of wind power systems. The performance evaluation, in general, is challenging due to the absence of a proper performance measure, the dependence of wind power generation on numerous factors, and the lack of knowledge about the effect of such factors. To address these challenges, we (i) developed a performance benchmark of wind power systems and defined a proper efficiency metric; (ii) devised a method to control for the effect of environmental factors that enables a fair comparison of the performance; (iii) established a general framework modeling adverse effects and characterized the wake effect.

In Chapter II and III, we evaluated the performance of wind power systems in terms of an efficiency metric. In other words, each power output is evaluated by comparing itself to the benchmark performance for the given wind input. Such an evaluation takes the effect of varying input conditions on the power output into consideration. This is essential for an evaluation of most renewable energy systems in which input is not controllable. From a practical point of view, the efficiency metric is especially important for the reliability assessment purpose as the metric is capable of providing insight into the inherent, yet never known, status of the system health. The additional treatment controlling for other environmental effects, described in Chapter III, enables better identification of the system

health trajectory, improving the reliability assessment of wind power systems. In fact, such a treatment controlling for other factors is useful whenever system performance needs to be compared. On the other hand, the analysis in Chapter IV can help mitigate the adverse wake effect by providing valuable information for the wind farm layout planning, wind turbine design, and operations of turbines and farms. With the general framework developed in Chapter IV, the research can be easily extended to the analysis of other adverse effects.

In addition to the benefits to wind energy applications, the methods developed in this dissertation also contribute to methodology development. The stochastic estimator satisfying the regular ultra-passum (RUP) law, developed in Chapter II, contributes to the field of production economics where the RUP is relevant. For some applications, in which the full set of inputs is difficult to identify or an influential input is not observable, noise modeling is essential. Chapter II demonstrates that with a significant level of noise, the proposed estimator outperforms the existing deterministic estimators that assume data as noise-free, suggesting its potential effectiveness for such an application. Furthermore, the spline based model developed in Chapter IV can be applied to any functional estimation problem that requires non-negativity (and smoothness) of the underlying function. A smoothing spline-based estimator imposing non-negativity is widely used for that purpose, but it is limited to the case of a single predictor only. In fact, the proposed estimator is a high-dimensional generalization of the smoothing spline-based estimator. Albeit possibly limited due to the curse of dimensionality and the scalability issues, the estimator sheds light on functional estimation that requires non-negative response.

V.2 Future work

In this dissertation, we have considered three sets of factors in modeling wind power output: (i) wind input; (ii) environmental variables; and (iii) adverse factors such as wake

effect. We believe these factors altogether influence the power output, but we have not yet defined a unified model including all the factors. In fact, we have conducted some preliminary studies predicting wind power output by using a model that includes wind input and environmental variables. However, none of the well-known methods in the statistics and machine learning fields could perform better than the estimator developed by Lee et al. (2015b) that assumes a specific additive multiplicative structure inspired by some physical observation. Different from other methods, the methods proposed in this dissertation impose a stronger physical knowledge such as S-shape for the power curve and non-negativity for the wake effect, and their effectiveness is well demonstrated throughout this dissertation. Therefore, it would be interesting to model the power output by integrating all the factors while preserving the stronger restrictions, which may deliver a better and more accurate wind power prediction model.

The application presented in Chapter III that traces annual efficiency of wind turbines is useful for understanding a turbine's aging process so beneficial for the reliability purpose. However, it may not be sufficient for short-term maintenance planning and health monitoring of wind turbines as some turbine components or subsystems can fail more frequently, for example, more than once a year for the electrical system (Sheng, 2013). For effective diagnosis and prognosis of such a failure occurring from time to time, it is important to monitor the turbine health in a real-time basis or at least in a short time period basis. The subset based analysis induced by the covariate density matching is not so practical in this case as the health monitoring deals with every single observation or an aggregated value over a short time interval. As such, the efficiency metric proposed in this dissertation needs to be improved so that it is completely (at least nearly) free of any other factor and hence does not require the subset based analysis.

REFERENCES

[1] Abadie, J. and Carpentier, J. (1969). Generalization of the Wolfe reduced gradient method to the case of nonlinear constraints. In: *Optimization*. Ed. by R. Fletcher. New York: Academic Press, p. 37–47.

[2] Abdullah, M. A., Yatim, A. H. M., Tan, C. W., and Saidur, R. (2012). A review of maximum power point tracking algorithms for wind energy systems. *Renewable and Sustainable Energy Reviews* 16 (5): 3220–3227.

[3] Adaramola, M. S. and Krogstad, P.-Å. (2011). Experimental investigation of wake effects on wind turbine performance. *Renewable Energy* 36 (8): 2078–2086.

[4] Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review* 8 (1): 67–77.

[5] Aigner, D., Lovell, C. A. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6 (1): 21–37.

[6] Ammara, I., Leclerc, C., and Masson, C. (2002). A viscous three-dimensional differential / actuator-disk method for the aerodynamic analysis of wind farms. *Journal of Solar Energy Engineering* 124 (4): 345–356.

[7] AWEA (2016). *U.S. number one in the world in wind energy production*. `http://www.awea.org/MediaCenter/pressrelease.aspx?ItemNumber=8463`. Accessed: 2016-11-03.

[8] Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30 (9): 1078–1092.

[9] Banker, R. D. and Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis* 3 (4): 401–415.

[10] Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* 22 (4): 469–483.

[11] Barthelmie, R. J. and Jensen, L. E. (2010). Evaluation of wind farm efficiency and wind turbine wakes at the Nysted offshore wind farm. *Wind Energy* 13 (6): 573–586.

[12] Barthelmie, R. J., Pryor, S. C., Frandsen, S. T., Hansen, K. S., Schepers, J. G., Rados, K., Schlez, W., Neubert, A., Jensen, L. E., and Neckelmann, S. (2010). Quantifying the impact of wind turbine wakes on power output at offshore wind farms. *Journal of Atmospheric and Oceanic Technology* 27 (8): 1302–1317.

[13] Betz, A. (1966). *Introduction to the Theory of Flow Machines*. 1st ed. Oxford: Pergamon Press.

[14] Bhutta, M. M. A., Hayat, N., Farooq, A. U., Ali, Z., Jamil, S. R., and Hussain, Z. (2012). Vertical axis wind turbine–A review of various configurations and design techniques. *Renewable and Sustainable Energy Reviews* 16 (4): 1926–1939.

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

[16] Boukhezzar, B., Siguerdidjane, H., and Maureen Hand, M. (2006). Nonlinear control of variable-speed wind turbines for generator torque limiting and power optimization. *ASME Journal of Solar Energy Engineering* 128 (4): 516–530.

[17] Carvalho, A., Gonzalez, M. C., Costa, P., and Martins, A. (2009). Issues on performance of wind systems derived from exploitation data. In: *Proc. Industrial Electronics, 2009. IECON'09. 35th Annual Conference of IEEE*. Porto, Portugal, p. 3599–3604.

[18] Chambers, R. G. (1988). *Applied Production Analysis: a Dual Approach*. Cambridge, England: Cambridge University Press.

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

[20] Deprins, D., Simar, L., and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In: *The Performance of Public Enterprises: Concept and Measurement*. Ed. by M. Marchand, P. Pestieau, and H. Tulkens. Amsterdam: North-Holland, p. 243–267.

[21] Diewert, W. E., Avriel, M., and Zang, I. (1981). Nine kinds of quasiconcavity and concavity. *Journal of Economic Theory* 25 (3): 397–420.

[22] DOE (2015). *Wind Vision: A New Era for Wind Power in the United States*. Tech. rep. Washington DC: U.S. Deparment of Energy.

[23] Drud, A. S. (1985). CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming* 31 (2): 153–191.

[24] — (1994). CONOPT—a large-scale GRG code. *ORSA Journal on Computing* 6 (2): 207–216.

[25] Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: *Constructive Theory of Functions of Several Variables*. Ed. by W. Schempp and K. Zeller. Heidelberg: Springer-Verlag, p. 85–100.

[26] Duckworth, A. and Barthelmie, R. J. (2008). Investigation and validation of wind turbine wake models. *Wind Engineering* 32 (5): 459–475.

[27] Emami, A. and Noghreh, P. (2010). New approach on optimization in placement of wind turbines within wind farm by genetic algorithms. *Renewable Energy* 35 (7): 1559–1564.

[28] Energy Information Administration (2016). *Word Pro - S7 - US Energy Information Administration*. `https://www.eia.gov/totalenergy/data/monthly/pdf/sec7_5.pdf`. Accessed: 2016-11-03.

[29] Eriksson, S., Bernhoff, H., and Leijon, M. (2008). Evaluation of different turbine concepts for wind power. *Renewable and Sustainable Energy Reviews* 12 (5): 1419–1434.

[30] Färe, R. and Shephard, R. W. (1977). Ray-homothetic production functions. *Econometrica: Journal of the Econometric Society* 45 (1): 133–146.

[31] Fenchel, W. and Blackett, D. W. (1953). *Convex Cones, Sets, and Functions*. Princeton: Princeton University, Department of Mathematics, Logistics Research Project.

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

[33] Frisch, R. (1964). *Theory of Production*. Dordrecht: Springer Science & Business Media.

[34] Gebraad, P. M. O., Teeuwisse, F. W., Wingerden, J. W., Fleming, P. A., Ruben, S. D., Marden, J. R., and Pao, L. Y. (2016). Wind plant power optimization through yaw control using a parametric model for wake effects—a CFD simulation study. *Wind Energy* 19 (1): 95–114.

[35] Gill, S., Stephen, B., and Galloway, S. (2012). Wind turbine condition assessment through power curve copula modeling. *IEEE Transactions on Sustainable Energy* 3 (1): 94–101.

[36] Hackman, S. T. (2008). *Production Economics: Integrating the Microeconomic and Engineering Perspectives*. Heidelberg: Springer-Verlag.

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

[38] Hansen, K. S., Barthelmie, R. J., Jensen, L. E., and Sommer, A. (2012). The impact of turbulence intensity and atmospheric stability on power deficits due to wind turbine wakes at Horns Rev wind farm. *Wind Energy* 15 (1): 183–196.

[39] Hanson, D. L. and Pledger, G. (1976). Consistency in concave regression. *The Annals of Statistics* 4 (6): 1038–1050.

[40] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. 2nd ed. New York: Springer-Verlag.

[41] Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Vol. 43. London: Chapman & Hall/CRC.

[42] Hayes, B. P., Ilie, I., Porpodas, A., Djokic, S. Z., and Chicco, G. (2011). Equivalent power curve model of a wind farm based on field measurement data. In: *Proc. IEEE PES Trondheim PowerTech 2011*. IEEE. Trondheim, Norway, p. 1–7.

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

[44] Holloway, C. A. (1979). Technical note—on the estimation of convex functions. *Operations Research* 27 (2): 401–407.

[45] Homola, M. C., Virk, M. S., Nicklasson, P. J., and Sundsbø, P. A. (2012). Performance losses due to ice accretion for a 5 MW wind turbine. *Wind Energy* 15 (3): 379–389.

[46] IEC12.1 (2005). *IEC 61400-12-1 Ed 1, Wind Turbines - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines*. Geneva, Switzerland: International Electrotechnical Commission.

[47] IEC12.2 (2013). *IEC 61400-12-2, Wind Turbines - Part 12-2: Power Performance of Electricity Producing Wind Turbines Based on Nacelle Anemometry*. Geneva, Switzerland: International Electrotechnical Commission.

[48] Jensen, N. O. (1983). *A note on wind generator interaction*. Tech. rep. Risø-M; No. 2411. `http://orbit.dtu.dk/files/55857682/ris_m_2411.pdf`. Roskilde, Denmark: Risø National Laboratory.

[49] Khalfallah, M. G. and Koliub, A. M. (2007). Effect of dust on the performance of wind turbines. *Desalination* 209 (1): 209–220.

[50] Khalid, M. and Savkin, A. V. (2012). A method for short-term wind power prediction with multiple observation points. *IEEE Transactions on Power Systems* 27 (2): 579–586.

[51] Krogstad, P.-Å. and Lund, J. A. (2012). An experimental and numerical study of the performance of a model turbine. *Wind Energy* 15 (3): 443–457.

[52] Kuosmanen, T. (2001). DEA with efficiency classification preserving conditional convexity. *European Journal of Operational Research* 132 (2): 326–342.

[53] — (2008). Representation theorem for convex nonparametric least squares. *The Econometrics Journal* 11 (2): 308–325.

[54] Kuosmanen, T. and 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.

[55] Laan, M. P., Sørensen, N. N., Réthoré, P.-E., Mann, J., Kelly, M. C., Troldborg, N., Schepers, J. G., and Machefaux, E. (2015). An improved k-ϵ model applied to a wind turbine wake in atmospheric turbulence. *Wind Energy* 18 (5): 889–907.

[56] Lee, C.-Y., Johnson, A. L., Moreno-Centeno, E., and Kuosmanen, T. (2013). A more efficient algorithm for convex nonparametric least squares. *European Journal of Operational Research* 227 (2): 391–400.

[57] Lee, G., Ding, Y., Xie, L., and Genton, M. G. (2015a). A kernel plus method for quantifying wind turbine performance upgrades. *Wind Energy* 18 (7): 1207–1219.

[58] Lee, G., Ding, Y., Genton, M. G., and Xie, L. (2015b). Power curve estimation with multivariate environmental factors for inland and offshore wind farms. *Journal of the American Statistical Association* 110 (509): 56–67.

[59] Lim, E. and Glynn, P. W. (2012). Consistency of multidimensional convex regression. *Operations Research* 60 (1): 196–208.

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

[61] Márquez, F. P. G., Tobias, A. M., Pérez, J. M. P., and Papaelias, M. (2012). Condition monitoring of wind turbines: techniques and methods. *Renewable Energy* 46: 169–178.

[62] McKay, P., Carriveau, R., and Ting, D. S.-K. (2013). Wake impacts on downstream wind turbine performance and yaw alignment. *Wind Energy* 16 (2): 221–234.

[63] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications* 9 (1): 141–142.

[64] Olesen, O. B. and Petersen, N. C. (2013). Imposing the Regular Ultra Passum law in DEA models. *Omega* 41 (1): 16–27.

[65] Olesen, O. B. and Ruggiero, J. (2014). Maintaining the Regular Ultra Passum Law in data envelopment analysis. *European Journal of Operational Research* 235 (3): 798–809.

[66] Papatheou, E., Dervilis, N., Maguire, E., Antoniadou, I., and Worden, K. (2015). A performance monitoring approach for the novel Lillgrund offshore wind farm. *IEEE Transactions on Industrial Electronics* 62 (10): 6636–6644.

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

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

[69] Pieralli, S., Ritter, M., and Odening, M. (2015). Efficiency of wind power production and its determinants. *Energy* 90 (1): 429–438.

[70] Prospathopoulos, J. M., Politis, E. S., Rados, K. G., and Chaviaropoulos, P. K. (2011). Evaluation of the effects of turbulence model enhancements on wind turbine wake predictions. *Wind Energy* 14 (2): 285–300.

[71] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd ed. New York: Springer-Verlag.

[72] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.

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

[74] Sanderse, B., Pijl, S. P., and Koren, B. (2011). Review of computational fluid dynamics for wind turbine wake aerodynamics. *Wind Energy* 14 (7): 799–819.

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

[76] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3): 683–690.

[77] Sheng, S. (2013). *Report on wind turbine subsystem reliability—a survey of various databases*. Tech. rep. NREL/PR-5000-59111. Golden, CO: National Renewable Energy Laboratory (NREL).

[78] Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise? *Journal of Productivity Analysis* 28 (3): 183–201.

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

[80] Statista (2016). *Wind power wholesale power prices United States 2015 | Statistics*. https://www.statista.com/statistics/217841/us-cumulative-capacity-weighted-average-wind-power-price/. Accessed: 2016-12-28.

[81] Troldborg, N., Larsen, G. C., Madsen, H. A., Hansen, K. S., Sørensen, J. N., and Mikkelsen, R. (2011). Numerical simulations of wake interaction between two wind turbines at various inflow conditions. *Wind Energy* 14 (7): 859–876.

[82] Uluyol, O., Parthasarathy, G., Foslien, W., and Kim, K. (2011). Power curve analytic for wind turbine performance monitoring and prognostics. In: *Proc. Annual Conference of the Prognostics and Health Management Society*. Vol. 2. Montreal, Canada, p. 1–8.

[83] Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica* 50 (4): 945–973.

[84] WAsP (2016). *Wind resources for energy production of wind turbines - WAsP*. http://www.wasp.dk/wasp. Accessed: 2016-12-11.

[85] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* 26 (4): 359–372.

[86] Wikipedia (2017). *Capacity factor — Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/wiki/Capacity_factor. Accessed: 2017-01-19.

[87] Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (1): 95–114.

[88] You, M., Byon, E., Jin, J., and Lee, G. (2017). When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms. *IISE Transactions* 49 (1): 84–95.

APPENDIX A

PROOFS

A.1 Proof of Theorem 1

We prove the theorem based on the result from Hanson and Pledger (1976) proving consistency of the Hildreth type estimator (Hildreth, 1954). We impose the same set of assumptions as Hanson and Pledger (1976) do; the assumptions are stated below.

Let $T \subset \Re$ be a subinterval of the real line having positive length and T_1 and T_2 partition T such that $t_1 \leq t_2$ whenever $t_1 \in T_1$ and $t_2 \in T_2$. Consider the ordered input observations $x_1, x_2, \ldots, x_n$ as a sequence of points in T. Correspondingly, the output observations $y_1, y_2, \ldots, y_n$ forms an independent sequence of random variables of which distribution is characterized by F_{x_k}. For each subset A_a of T_a for $a = 1, 2$, define

$$N_n(A_a) = \sum_{k=1}^{n} I_{A_a}(x_k) = \#\{k : 1 \leq k \leq n,\ x_k \in A_a\}$$

where $\#\{S\}$ counts the number of elements in a set S. Also, for $y \geq 0$, define

$$G_a(y) = \sup_{x \in T_a}\{F_x[\psi(x) - y] + 1 - F_x[\psi(x) + y]\}$$

for $a = 1, 2$. Then, G_a provides a uniform bound on the tails of the distributions of the error random variables $y_k - \psi(x_k)$, i.e.,

$$P\{|y_k - \psi(x_k)| \geq y\} \leq G_a(y),\ \forall k \text{ s.t. } x_k \in T_a.$$

Based on these notations and definitions, we assume the followings.

(A_1) $\psi(x)$ is continuous on T, convex on T_1, and concave on T_2.

(A_2) For each subinterval I_a of T_a for $a = 1, 2$ having positive length, lim inf $N_n(I_a)/n > 0$.

(A_3) $\lim_{y\to\infty} G_a(y) = 0$ for $a = 1, 2$.

(A_4) $\int_0^\infty y^2 |dG_a(y)| = Q_a^2 < \infty$ for $a = 1, 2$.

Then, Theorem 1 is a direct consequence of Hanson and Pledger (1976) and the following lemmas.

Lemma 2. *Suppose that inflection point x^* is given, so the convex and concave regions can be determined accordingly. Among the ordered input observations, let the last observation in the convex region and the first observation in the concave region be $x_{n_{cvx}}$ and $x_{n_{cvx}+1}$, respectively. Also, let $\psi(x)$ be differentiable up to the second order. If $x_{n_{cvx}+1} - x_{n_{cvx}} \to 0$ as $n \to \infty$, the least square estimator defined by* (II.10) *is a consistent estimator of $\psi(x)$.*

Proof. The optimization problem in (II.10) integrates two optimization problems, one for convex region and another for concave region, with some restriction relating the two. Had we considered the convex and the concave regions separately, the independent CNLS problems for each region would have been

$$
\begin{aligned}
\min_{\boldsymbol{\beta}} \quad & \sum_{k=1}^{n_{cvx}} (y_k - \hat{y}_k)^2 \\
\text{s.t.} \quad & \hat{y}_k = \beta_{k,0} + \beta_{k,1} x_k, \ \forall k = 1, \ldots, n_{cvx}, \\
& \beta_{k,1} = \frac{\hat{y}_{k+1} - \hat{y}_k}{x_{k+1} - x_k}, \ \forall k = 1, \ldots, n_{cvx} - 1, \\
& \beta_{k,1} \le \beta_{k+1,1}, \ \forall k = 1, \ldots, n_{cvx} - 2, \\
& \beta_{k,1} \ge 0, \ \forall k = 1, \ldots, n_{cvx} - 1, \\
& \hat{y}_k \ge 0, \ \forall k = 1, \ldots, n_{cvx},
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
\min_{\boldsymbol{\beta}} \quad & \sum_{k=n_{cvx}+1}^{n} (y_k - \hat{y}_k)^2 \\
\text{s.t.} \quad & \hat{y}_k = \beta_{k,0} + \beta_{k,1} x_k, \ \forall k = n_{cvx}+1, \ldots, n, \\
& \beta_{k,1} = \frac{\hat{y}_k - \hat{y}_{k-1}}{x_k - x_{k-1}}, \ \forall k = n_{cvx}+2, \ldots, n, \\
& \beta_{k-1,1} \geq \beta_{k,1}, \ \forall k = n_{cvx}+3, \ldots, n, \\
& \beta_{k,1} \geq 0, \ \forall k = n_{cvx}+2, \ldots, n, \\
& \hat{y}_k \geq 0, \ \forall k = n_{cvx}+1, \ldots, n,
\end{aligned} \tag{A.2}
$$

respectively. The solutions from (A.1) and (A.2), respectively, estimate $\psi(x)$ defined over $[x_1, x_{n_{cvx}}]$ and $[x_{n_{cvx}+1}, x_n]$; let the estimates be $\hat{\psi}^{cvx}(x)$ and $\hat{\psi}^{ccv}(x)$. Problem (II.10) integrates (A.1) and (A.2) imposing an additional hyperplane connecting $(x_{n_{cvx}}, \hat{y}_{n_{cvx}})$ and $(x_{n_{cvx}+1}, \hat{y}_{n_{cvx}+1})$ of which slope

$$
\beta_{n_{cvx},1} = \beta_{n_{cvx}+1,1} = \frac{\hat{y}_{n_{cvx}+1} - \hat{y}_{n_{cvx}}}{x_{n_{cvx}+1} - x_{n_{cvx}}}
$$

is no less than $\beta_{n_{cvx}-1,1}$ and no greater than $\beta_{n_{cvx}+2,1}$.

By the consistency results from Hanson and Pledger (1976), $\hat{\psi}^{cvx}(x)$ is a consistent estimator of $\psi(x)$ in the interior of its support, i.e., $(x_1, x_{n_{cvx}})$. The integrated problem (II.10) restricts the convex problem (A.1) only for its support near the upper boundary. For sufficiently large n, the integrated problem takes advantage of the consistency of the least square estimator $\hat{\psi}^{cvx}(x)$, so there must be some x_l such that $x_k, \forall k = 1, \ldots, l$ have the same fitted values from the integrated problem and the convex problem. For $\hat{y}_k, k = l+1, \ldots, n_{cvx}$ taken from the two problems to be different, it is necessary to have $\beta_{l,1} = \beta_{n_{cvx},1}$ from the integrated problem, so that the third set of constraints in (A.1) are binding for these observations meaning that the integrated problem restricts the convex problem. Similarly, there must be an x_u such that $x_k, \forall k = u, \ldots, n$ have the same fitted values from

the integrated problem and the concave problem (A.2). Here, the necessary condition for $u > n_{cvx} + 1$ is $\beta_{n_{cvx}+1,1} = \beta_{u,1}$ in the integrated problem. This along with $\beta_{l,1} = \beta_{n_{cvx},1}$ implies that the $\hat{\psi}(x)$ should have a linear segment connecting x_l and x_u.

If $\psi'(x_l) = \psi'(x_u)$ and $\psi''(x) = 0$ for $\forall x \in [x_l, x_u]$, we obtain the desired result. Otherwise, $\psi(x)$ still reveals convexity followed by concavity over the interval $[x_l, x_u]$. As n grows, the least square estimator in the integrated problem requires more hyperplanes to fit the S-shape curve in $[x_l, x_u]$ yielding $\beta_{l,1} < \beta_{n_{cvx},1}$ and $\beta_{n_{cvx}+1,1} > \beta_{u,1}$. Since $\beta_{l,1}$ and $\beta_{u,1}$ are no longer restricted by $\beta_{n_{cvx},1}$ and $\beta_{n_{cvx}+1,1}$, respectively, x_l increases and x_u decreases, shrinking the interval $[x_l, x_u]$. By repeating this procedure and by induction,

$$\hat{\psi}(x_k) = \hat{\psi}^{cvx}(x_k) \xrightarrow{p} \psi(x_k),\ k = 1, \ldots, n_{cvx} - 1$$
$$\hat{\psi}(x_k) = \hat{\psi}^{ccv}(x_k) \xrightarrow{p} \psi(x_k),\ k = n_{cvx} + 2, \ldots, n.$$

For the boundary points $x_{n_{cvx}}$ and $x_{n_{cvx}+1}$, $\hat{\psi}^{cvx}(x_{n_{cvx}})$ and $\hat{\psi}^{ccv}(x_{n_{cvx}+1})$ are not guaranteed to be a consistent estimator of $\psi(x_{n_{cvx}})$ and $\psi(x_{n_{cvx}+1})$, respectively. However, by assumption, $x_{n_{cvx}+1} - x_{n_{cvx}} \to 0$ also implying $x_{n_{cvx}+2} - x_{n_{cvx}-1} \to 0$; equivalently, $x_{n_{cvx}+2} \downarrow x^*$ and $x_{n_{cvx}-1} \uparrow x^*$. Then, continuity of $\psi(x)$ yields

$$\hat{\psi}(x_{n_{cvx}-1}) \xrightarrow{p} \psi(x_{n_{cvx}-1}) \to \psi(x^*)$$
$$\hat{\psi}(x_{n_{cvx}+2}) \xrightarrow{p} \psi(x_{n_{cvx}+2}) \to \psi(x^*).$$

By the imposed continuity of $\hat{\psi}(x)$, $\hat{\psi}(x_{n_{cvx}}) \xrightarrow{p} \psi(x^*)$ and $\hat{\psi}(x_{n_{cvx}+1}) \xrightarrow{p} \psi(x^*)$, concluding the proof □

Lemma 3. *For given n, let* $x^*_- = max\{x_k : x_k < x^*,\ k = 1, \ldots, n\}$ *and* $x^*_+ = min\{x_k : x_k > x^*,\ k = 1, \ldots, n\}$. *Suppose that the sequences of* $\{x^*_-\}_n$ *and* $\{x^*_+\}_n$ *both approach* x^* *as* $n \to \infty$. *Also, let the number of grid points tend to infinity in the way that there is at*

least one grid point between x^*_- and x^*_+ while $n \to \infty$. Then, the grid point yielding the least objective function value in (II.10) *approaches the true inflection point, i.e., $\hat{x}^* \to x^*$.*

Proof. Lemma 2 states that as $n \to \infty$, the estimator minimizing the sum of squared errors (SSE; the objective function value defined in (II.10)) with a correctly specified inflection point is a consistent estimator of $\psi(x)$. So, if there is any grid point that derives a consistent estimator, the corresponding optimization problem should minimize SSE. By assumption, we always have a grid point defining the same convex and concave regions as the true inflection point does, and the grid point should be the one minimizing SSE for sufficiently large n. Such a grid point is our inflection point estimate $\hat{x}^*$, and we have $x_{n_{cvx}} < \hat{x}^* < x_{n_{cvx}+1}$ where $x_{n_{cvx}} = x^*_-$ and $x_{n_{cvx}+1} = x^*_+$. By assumption, x^*_- and x^*_+ converge to x^* as n grows, so $\hat{x}^* \to x^*$ □

APPENDIX B

SUPPLEMENTAL MATERIALS FOR CHAPTER II

B.1 Determination of the initial solution and the bounds of decision variables for base isoquant estimation

When performing a nonlinear optimization, a proper initial solution as well as appropriate restriction on the lower and upper bounds of unknown variables is critical for the convergence to an optimal solution. A simple and good initial solution for (II.7) can be achieved by applying multivariate linear regression for which one of the inputs plays a role of response while others remain as regressors. For example, consistent with the role of inputs in (II.7b), if we set the pth input as a response variable,

$$x_{ip} = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \ldots + \delta_{p-1} x_{i(p-1)} + \epsilon_i,$$

for $i \in \mathcal{I}$ where δ_0 and δ_q for $q = 1, \ldots, p-1$ are the coefficients of an affine hyperplane and ϵ_i represents a random error. Once the model is fitted, we consider a radial projection of input vectors onto this fitted hyperplane so that $\hat{\gamma}_i \boldsymbol{X}_i$ for $i \in \mathcal{I}$ is on the hyperplane. Given the coefficient estimates $\hat{\delta}_0$ and $\hat{\delta}_q$'s, we have

$$\hat{\gamma}_i x_{ip} = \hat{\delta}_0 + \hat{\delta}_1 \hat{\gamma}_i x_{i1} + \hat{\delta}_2 \hat{\gamma}_i x_{i2} + \ldots + \hat{\delta}_{p-1} \hat{\gamma}_i x_{i(p-1)},$$

which reduces to

$$\hat{\gamma}_i = \hat{\delta}_0 / (x_{ip} - \hat{\delta}_1 x_{i1} - \ldots - \hat{\delta}_{p-1} x_{i(p-1)}).$$

For the initial values of α_i and β_{iq}, we can use the coefficient estimates directly, i.e., $\alpha_i = \hat{\delta}_0$ for $\forall i \in \mathcal{I}$ and $\beta_{iq} = \hat{\delta}_q$ for $\forall i \in \mathcal{I}$, $\forall q = 1, \ldots, p-1$. The hyperplane defined by

the coefficient estimates is affine, so it is convex. With a properly selected subset $L(y^*)_\alpha$, the hyperplane should be monotone decreasing with respect to the predictors, i.e., $\hat{\delta}_q < 0$ for $\forall q = 1, \ldots, p-1$. The values for $\hat{\gamma}_i$, α_i, and β_{iq} derived above then provide a feasible initial solution to (II.7).

In addition, we explicitly set the bounds of $\hat{\gamma}_i$, which will also restrict the bounds of α_i and β_{iq} through the relationship defined in (II.7b). By the definition of $\hat{\boldsymbol{X}}_i$, we have $\hat{\boldsymbol{X}}_i = \hat{\gamma}_i \boldsymbol{X}_i$, and this can be rewritten as $\hat{\gamma}_i = ||\hat{\boldsymbol{X}}_i||/||\boldsymbol{X}_i||$ where $||\cdot||$ calculates the Euclidean norm. With such an expression,

$$\min_i\{||\hat{\boldsymbol{X}}_i||\}/\max_i\{||\boldsymbol{X}_i||\} \leq \hat{\gamma}_i \leq \max_i\{||\hat{\boldsymbol{X}}_i||\}/\min_i\{||\boldsymbol{X}_i||\}$$

for $\forall \hat{\gamma}_i$ for $i \in \mathcal{I}$. Since $\hat{\boldsymbol{X}}_i$ is unknown *a priori*, the numerators in the lower and the upper bounds need to be substituted with some known quantities. By the convexity constraints and the minimization of radial deviation in (II.7), $||\hat{\boldsymbol{X}}_i||$ for any $i \in \mathcal{I}$ cannot exceed $\max_i\{||\boldsymbol{X}_i||\}$ except those defining the boundary of the isoquant estimate. On the other hand, some $||\hat{\boldsymbol{X}}_i||$ can be smaller than $\min_i\{||\boldsymbol{X}_i||\}$; see Figure B.1.

In Figure B.1, suppose that the point A has the smallest Euclidean norm so that no other observed input vector is contained in the quadrant of which radius equals $||A||$. Depending on the locations of other input vectors, the norm of the radial projection of A could be smaller than that of the point A, i.e., $||A'|| < ||A||$ as in Figure B.1(a). For the most extreme case, if the norm of some points on the isoquant estimate is the same with $||A||$, $||A'||$ can achieve its minimum when the points on the isoquant estimate has zero input for either input as shown in Figure B.1(b). In this case, $||A'|| = ||A||/\sqrt{2}$, and this can be generalized by $||A'|| = ||A||/\sqrt{p}$ in p-dimensional input space. Therefore, we have $\min_i\{||\hat{\boldsymbol{X}}_i||\} \geq \min_i\{||\boldsymbol{X}_i||\}/\sqrt{p}$. Based on these observations, we restrict $\hat{\gamma}_i$ for $\forall i \in \mathcal{I}$ to

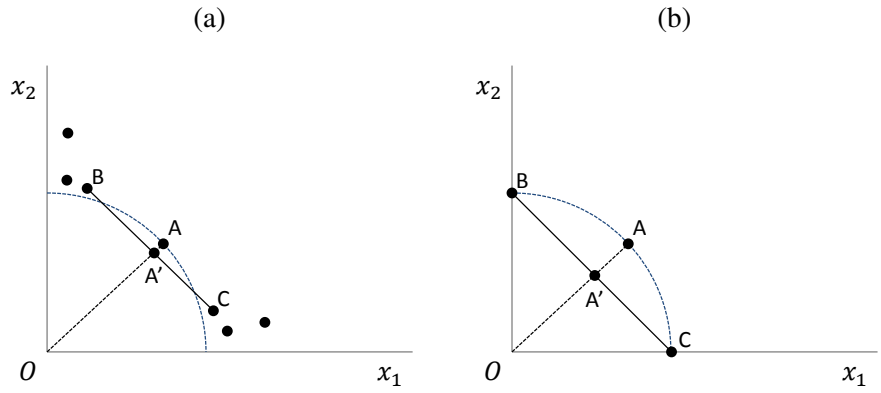

Figure B.1: Relationship between an input vector with the minimum Euclidean norm and its projection onto the isoquant estimate. The point A has the smallest Euclidean norm among the observed input vectors, and the dotted quadrant illustrates the set of points that has the same norm with the point A. Point B and C are some nearby points on the isoquant estimate, and point A' represents the radial projection of A onto the isoquant estimate.

be

$$\min_i\{||\boldsymbol{X}_i||\}/(\sqrt{p}\cdot\max_i\{||\boldsymbol{X}_i||\}) \leq \hat{\gamma}_i \leq \max_i\{||\boldsymbol{X}_i||\}/\min_i\{||\boldsymbol{X}_i||\}.$$

B.2 Consistency of the isoquant estimator under multiplicative and radial random errors

In this section, we demonstrate consistency of the isoquant estimator defined in (II.7) under multiplicative and radial random errors, by showing consistency of $\hat{\gamma}_i$ numerically via Monte Carlo simulation.

For the simulation, we generate data in a way similar to the DGP described in Section II.5. However, different from the DGP, here we consider data of which output value is the same for all, so that they altogether can define an isoquant. Recall that we have $y = \phi(x_1, x_2) = F\left(g(x_1, x_2)\right)$, where $F(z) = 15/(1 + \exp(-5\ln z))$. Without loss of generality, we will use the output value of the inflection point of F to define an isoquant. With a little algebra, the inflection point of F can be derived as $(z, F(z)) = ((2/3)^{1/5}, 6)$.

Then, we define an isoquant associated with $y = 6$ by finding a set of input vectors $\boldsymbol{X} = (x_1, x_2)$ satisfying $g(x_1, x_2) = (2/3)^{1/5}$.

We first generate angles $\eta \sim Unif(0.05, \pi/2 - 0.05)$, and for each given η calculate modulus ω as follows:

$$\left(\frac{2}{3}\right)^{\frac{1}{5}} \left\{\beta(\cos\eta)^{\frac{\sigma-1}{\sigma}} + (1-\beta)(\sin\eta)^{\frac{\sigma-1}{\sigma}}\right\}^{\frac{-\sigma}{\sigma-1}},$$

where $\beta = 0.45$ and $\sigma = 1.51$. Let $\tilde{\boldsymbol{X}} = (\omega\cos\eta,\ \omega\sin\eta)$ denote the set of points defining the isoquant associated with $y = 6$. To generate multiplicative and radial random noise, we first draw v from a normal distribution with mean 0 and variance σ_v^2, i.e., $v \sim N(0, \sigma_v^2)$, and calculate the simulated input $\boldsymbol{X} = \tilde{\boldsymbol{X}} \cdot \exp(v)$.

To show the convergence of $\hat{\gamma}_i$ to the true value γ_i with an increasing sample size n, we generate $\tilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ for $n = 30, 50, 100, 200, 300$ for given σ_v. Similar to the simulation experiments in Section II.5, we investigate the effect of noise level by considering different values of σ_v of 0.05, 0.1, 0.15. The comparison is again based on 90% confidence intervals (CIs) of RMSE obtained by 100 bootstrap resampling. The RMSE for this experiment is defined by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\gamma_i - \hat{\gamma}_i)^2},$$

where $\gamma_i = ||\tilde{\boldsymbol{X}}_i||/||\boldsymbol{X}_i||$ and $\hat{\gamma}_i = ||\hat{\boldsymbol{X}}_i||/||\boldsymbol{X}_i||$.

Figure B.2 presents the 90% CIs changing with different sample sizes. Obviously, there is an overall decreasing trend of the CIs and the means, and the values get closer to zero as the sample size increases although the convergence rate gets slower for a higher noise level.

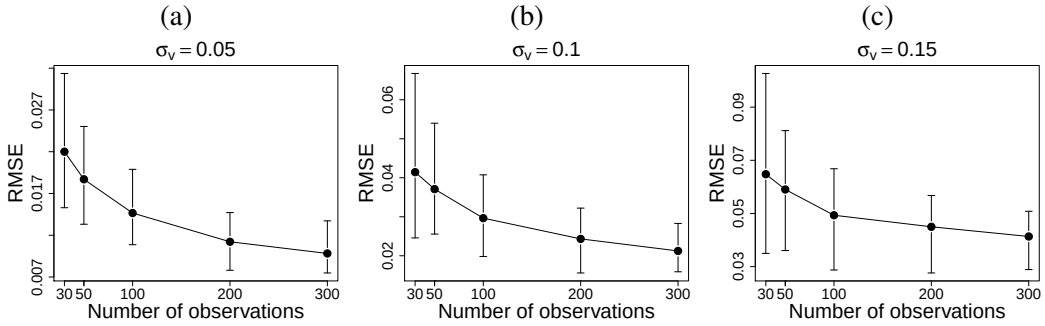

Figure B.2: Decreasing RMSE with increasing sample sizes. The bars represents the 90% CIs of RMSE, and the solid dots indicate the means of RMSE for each given sample size.

B.3 Calculation of base output selection metrics

In this section, we define two metrics of angle_k and unif_k originally suggested by O&R for the base output selection. For simplicity, suppose that we have two-dimensional inputs. Then, any input vector $\boldsymbol{X}_i \in L(y^*)_\alpha$ can be expressed in terms of polar coordinates (ω_i, η_i) where ω_i and η_i denote the radial coordinate and the angular coordinate, respectively. Based on the polar coordinate system, angle_k is calculated as

$$\text{angle}_k = \max\{\eta_i : \, i \in \mathcal{I}(k)\} - \min\{\eta_i : \, i \in \mathcal{I}(k)\},$$

measuring the maximum range of the polar angles. The notation $\mathcal{I}(k)$ represents dependency of $\mathcal{I}$ on k. To calculate unif_k, we first sort η_i's in the increasing order and define $\tilde{\eta}_i \in [0, 1]$ by rescaling η_i as

$$\tilde{\eta}_i = \frac{\eta_i - \min\{\eta_k : \, k = 1, \ldots, n\}}{\max\{\eta_k : \, k = 1, \ldots, n\} - \min\{\eta_k : \, k = 1, \ldots, n\}},$$

where the support of η_i is estimated by using all observations (indexed by k) rather than the subset $i \in \mathcal{I}(k)$. By assuming $\tilde{\eta}_i$ is uniformly distributed, its point-wise empirical cu-

mulative distribution function (cdf) and theoretical cdf can be calculated as $P_{\text{emp}}(\tilde{\eta}_i) = \tilde{\eta}_i$ and $P_{\text{unif}}(\tilde{\eta}_i) = i/n_\alpha$, respectively. Then, unif_k measures the absolute difference between the two cdfs within the support of $\tilde{\eta}_i$, so

$$\text{unif}_k = \int_0^1 |P_{\text{emp}}(\tilde{\eta}) - P_{\text{unif}}(\tilde{\eta})| \, d\tilde{\eta},$$

and is approximated by the trapezoidal rule for which we evaluate the integrand at $\tilde{\eta}_i$ for $i \in \mathcal{I}(k)$.

For high dimensional inputs, angle_k and unif_k can be calculated for each angular dimension in a similar manner. We suggest to aggregate the multiple measures for each angular dimension first by taking an average and then use the averaged measure for the comparison with a threshold; instead of applying the threshold directly to the multiple measures for each angular dimension.

B.4 Alternative base output selection and isoquant estimation procedure

In Section II.3.2, we introduce multiple criteria for the selection of base output. The usage of multiple criteria can improve the finite sample performance of an isoquant estimator. However, it may cause more computational burdens and may sometimes raise an argument about how to decide the weights for different criteria. Such a concern can be alleviated by using a single criterion which does not require to solve the optimization problem in (II.7) for every observed output value.

Among the three criteria defined in Section II.3.2, the tight bounds of an isoquant estimator is the most desirable property in our stochastic setting since it has a great potential to reduce the variance of the estimator. For the calculation of the tightness, we have used some deviation measured in the input space, but we may approximate it by considering some deviation in the output space. Recall from (II.5), we select a subset of data whose output value is within $[y_l, y_u]$, and we estimate an isoquant using this subset of data. $[y_l, y_u]$

then forms the bounds of the selected output y_i for $i \in \mathcal{I}$, and their tightness can be evaluated by the length of the interval, $y_u - y_l$. The length of the interval is strongly related to the output density estimate $\hat{\pi}(y)$ defined in (II.6); the larger the density estimate is, the tighter the bounds become. This suggests that we can use $\hat{\pi}(y)$ as a metric evaluating the tightness of the output bounds and approximating the tightness of the bounds of an isoquant estimator. An advantage of using $\hat{\pi}(y)$ over $y_u - y_l$ is that $\hat{\pi}(y)$ can be calculated without determining any subset of data, $(\boldsymbol{X}_i, y_i)$ for $i \in \mathcal{I}$, thus simplifying the calculation procedure. Then, by using $\hat{\pi}(y)$, we can determine the base output as $y^* = \text{argmax } \hat{\pi}(y)$, the mode of the density estimate.

If necessary, one may also replace the nonlinear programming in (II.7) with a simpler optimization model which can reduce the computational complexity caused by the non-linearity. For the simpler model, we minimize deviation of input vectors measured in a single input dimension rather than the radial deviation associated with all the input dimensions. Specifically, one may solve the following quadratic programming to estimate $\text{Isoq}L(y^*)_{\text{CM}}$

$$
\begin{aligned}
\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \quad & \sum_{i=1}^{n_\alpha} (x_{ip} - \hat{x}_{ip})^2 \\
\text{s.t.} \quad & \hat{x}_{ip} = \alpha_i + \sum_{q=1}^{p-1} \beta_{iq} x_{iq}, \ \forall i \in \mathcal{I}, \\
& \alpha_i + \sum_{q=1}^{p-1} \beta_{iq} x_{iq} \geq \alpha_j + \sum_{q=1}^{p-1} \beta_{jq} x_{iq}, \ \forall i, j \in \mathcal{I}, j \neq i, \\
& \beta_{iq} \leq 0, \ \forall i \in \mathcal{I}.
\end{aligned} \tag{B.1}
$$

In (B.1), we use notations similar to the ones used in (II.7) since the decision variables play similar roles in both optimization problems (although their values would be different). Also, as has been done in (II.7), we use the pth input as a response and the other

inputs as regressors when defining hyperplanes and impose the convexity and monotonicity of the isoquant estimator by restricting these hyperplanes. The distinction between the two optimization problems comes from how to define the fitted values and the objective function. In (B.1), the fitted values differ from the observed input vectors only for a single input dimension, and the objective function minimizes such deviation defined on the single dimension.

Due to the simplification and approximation, an isoquant estimator achieves better computational efficiency but likely poorer estimation performance. By using the O&R DGP described in Section II.5, we illustrate the performance difference between the alternatives and the estimator proposed in Section II.3. Note that we have two alternatives each for a different part of the entire isoquant estimation procedure; one for the base output selection and another for the isoquant estimation based on the subset determined by the selected base output. Through Monte Carlo simulation, we compare three estimators for which we use 1) single criterion and quadratic programming considering x_1 as a response ('QP X1'), 2) single criterion and quadratic programming considering x_2 as a response ('QP X2'), and 3) multiple criteria and nonlinear programming ('NLP'). The performance of the three estimators is evaluated by RMSE defined in (II.13).

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, α is relatively large (α=0.3) because we impose a minimum on the number of observations used for estimating the base isoquant. This can lead to y_l being less than zero. In fact, the O&R DGP generates 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 when using the single criterion $\hat{\pi}(y)$, we adjust (II.5) as follows: set $y_l = y^*/2$ and determine y_u such that $\int_{y_l}^{y_u} \hat{\pi}(y)dy = \alpha$. Still, 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

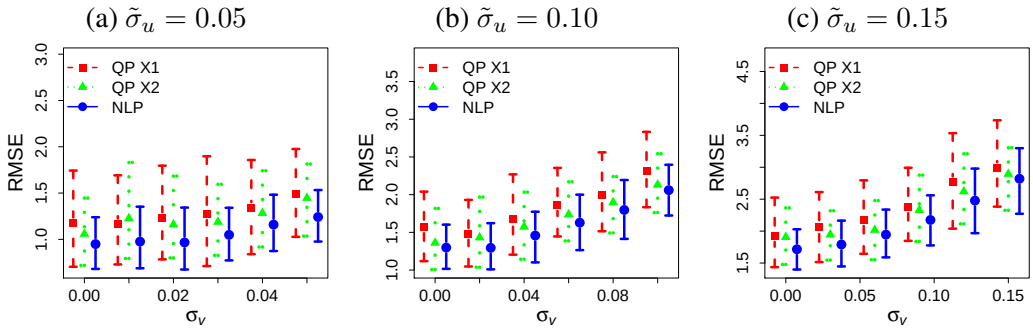

Figure B.3: Comparison of the alternative estimators and the original estimator.

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 B.3 shows the comparison results. Overall, the original isoquant estimator ('NLP') performs better than the alternative estimators, but the performance deterioration due to the approximation is not so pronounced relative to the case of comparing to the O&R estimator (see Figure II.5). In fact, the performance of the alternative estimators get closer to the performance of the original estimator as the magnitude of $\tilde{\sigma}_u$ and σ_v relative to the magnitude of $\phi(\boldsymbol{X})$ increases, so a user may consider using the alternatives if it is preferred.

One interesting phenomenon observed in Figure B.3 is that the estimator 'QP X2' in general outperforms 'QP X1' and its performance is often quite close to that of 'NLP'. This indicates that minimizing the deviation in terms of the x_2 dimension provides overall better results. This is most likely because the O&R DGP assigns a higher weight to x_2 than to x_1 when generating the output values. Recall that $g(x_1, x_2) = \left(\beta x_1^{(\sigma-1)/\sigma} + (1-\beta)x_2^{(\sigma-1)/\sigma}\right)^{\sigma/(\sigma-1)}$ where $\beta = 0.45$. Based on this knowledge, we recommend to apply OLS, i.e., $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, to determine the input dimension for which the deviation of input vectors is calculated; choose the input with a larger coefficient estimate. Under the O&R DGP, the resulting coefficient estimates averaged for each of 18 scenarios (averaged over

100 datasets in each scenario) range from 3.26 to 4.12 for $\hat{\beta}_1$ and from 4.35 to 5.25 for $\hat{\beta}_2$.

B.5 Imposing the weak essentiality axiom on the S-shaped averaged practice curve estimation

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 (II.10). To address the non-uniqueness issue, Kuosmanen (2008) has constructed an explicit representor function, so for the concave region,

$$\tilde{\psi}(x) = \min\left\{\hat{\beta}_{k,0} + \hat{\beta}_{k,1}x : \; k = n_{cvx}+1, \ldots, n\right\}.$$

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:

$$\tilde{\psi}_{\min}(x) = \min\{\beta_0 + \beta_1 x : \; \beta_0 + \beta_1 x_k \geq \hat{y}_k, \, \forall k = n_{cvx}, \ldots, n, \\ \beta_0 \in \Re, \, \beta_1 \in \Re_+\}. \quad \text{(B.2)}$$

In (B.2), the parameters β_0 and β_1 are reestimated, so they can be distinct from $\hat{\beta}_{k,0}$ and $\hat{\beta}_{k,1}$ obtained from (II.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 (B.2) and take the maximum instead of the minimum. Then, our estimator $\hat{\psi}(x)$ for the S-shaped production function is

$$\hat{\psi}(x) = \begin{cases} \max\{\beta_0 + \beta_1 x : \ \beta_0 + \beta_1 x_k \leq \hat{y}_k, \, \forall k = 1, \ldots, n_{cvx}, \\ \qquad \beta_0 \in \Re_-, \, \beta_1 \in \Re_+\}, & \text{if } x \leq x^* \\ \min\{\beta_0 + \beta_1 x : \ \beta_0 + \beta_1 x_k \geq \hat{y}_k, \, \forall k = n_{cvx} + 1, \ldots, n, \\ \qquad \beta_0 \in \Re, \, \beta_1 \in \Re_+\}, & \text{if } x \geq x^* \end{cases} \tag{B.3}$$

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

Proposition 4. *Production function estimate $\hat{\psi}$ passes through the origin if $\hat{\beta}_{k,0} \leq 0$ for $\forall k = 1, \ldots, n_{cvx}$ and $\hat{y}_k \geq 0$ for $\forall k = 1, \ldots, n$, in addition to other constraints in* (II.10).

Proof. We prove the statement considering two mutually exclusive cases: (i) $\hat{y}_1 = 0$ and (ii) $\hat{y}_1 > 0$. As the data set has been rearranged in non-decreasing order of the input, $\hat{y}_1$ represents the point-wise estimate of ψ for the minimum input level.

(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 $0 < x_1 < x^*$, otherwise x^* cannot be an inflection point estimate. Then,

$$\hat{\psi}(x_1) = \max\{\beta_0 + \beta_1 x_1 : \ \beta_0 + \beta_1 x_k \leq \hat{y}_k, \, \beta_0 \in \Re_-, \\ \beta_1 \in \Re_+, \, k = 1, \ldots, n_{cvx}\} = \hat{y}_1 = 0$$

which implies there exist $\beta_0 = \beta_1 = 0$ such that $\beta_0 + \beta_1 x_k \leq \hat{y}_k$ for $k = 1, \ldots, n_{cvx}$. So,

$$\hat{\psi}(0) = \max\{\beta_0 : \ \beta_0 + \beta_1 x_k \leq \hat{y}_k, \ \beta_0 \in \Re_-, \\ \beta_1 \in \Re_+, \ k = 1, \ldots, n_{cvx}\} = 0$$

because β_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 $k = 1, \ldots, n_{cvx}$. 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. For $k = 1, \ldots, n_{cvx}$, we can write

$$\hat{y}_k \geq \hat{y}_1 + \tilde{\beta}_{0,1}(x_k - x_1) = \hat{y}_1 - \tilde{\beta}_{0,1} x_1 + \tilde{\beta}_{0,1} x_k = 0 + \tilde{\beta}_{0,1} x_k.$$

Therefore, $\beta_0 = 0$ and $\beta_1 = \tilde{\beta}_{0,1}$ satisfies $\beta_0 + \beta_1 x_k \leq \hat{y}_k$ for $k = 1, \ldots, n_{cvx}$, so $\hat{\psi}(0) = 0$ □