NONPARAMETRIC SHAPE-CONSTRAINED MODELS FOR PRODUCTION ECONOMICS

A Dissertation

by

KEVIN PIERRE A. LAYER

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,	Andrew Johnson
Committee Members,	Erick Moreno-Centeno
	Mark Lawley
	Jianhua Huang
Head of Department,	Mark Lawley

December 2019

Major Subject: Industrial Engineering

ABSTRACT

In production economics, the cost function is a critical tool used to infer productivity and efficiency measures to describe the key features of an industry. This dissertation investigates non-parametric estimators with shape constraints. The goals are to improve researchers' understanding and illustrate the advantages of this set of estimators over other commonly used estimators.

First, the dissertation studies the direction selection for stochastic directional distance functions. Unlike much of the published literature on directional distance functions, the analysis is performed in stochastic settings. Applying a recently developed non-parametric shape-constrained estimator on a set of simulations, user guidelines about selecting the direction, a key tuning parameter for such estimators, are given. The estimator is tested and compared to other estimators by applying it to a cost function estimation. An application of stochastic directional distance function estimator to the US hospitals industry gives insights into the industry such as most productive scale size and output trade-off information.

Second, several approximations of shape-constrained non-parametric estimators are analyzed. The approximations consist of piece-wise linear versus smooth (at least of class $\mathbf{C}^2$) estimated functions and coordinate-wise versus global constraints. The fitting performance and the shape constraints violations percentages are the main criteria established for the comparison. New estimators are developed for the analysis, in particular a B-spline based shape-constrained estimator for smooth cases. Based on the results obtained on a range of simulations, guidelines are determined to help users pick the best estimator, among the ones considered in the study, depending on the characteristics of the data. Finally additional insights on the US hospital industry are provided while showcasing the implementation of some of the introduced estimators on real data.

DEDICATION

To my parents, Florence and Francis; to my sister, Tiphaine who is also on a path to be a doctor but the type that cures people.

ACKNOWLEDGMENTS

First and foremost I thank Dr. Johnson. You shared your passion and helped me find interesting problems to tackle, learning so much and developing my skills along the way. Your guidance and support throughout the process meant a lot for me. I would also like to express my gratitude to all the contributors, Dr. Ferrier and Dr. Sickles for being co-authors on papers and Dr. Huang for providing help and advice on the questions about spline estimators. I would also like to thank the committee members Dr. Moreno and Dr. Lawley for their insightful feedback.

I don't forget all the coworkers and friends who I spent so much time with through these years. Our interactions made it easier in the sometimes difficult day-to-day routine. In particular Jose-Luis as the first leader, trail blazer, of the research group when I joined, you had a very positive influence on Daisuke and I; and of course Daisuke with whom I spent four years, your energy, discipline and kindness only match your smartness, you are an inspiration and a friend.

Finally thank you for my family to have been patient and supportive for me during this long process.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Dr. Andrew L. Johnson [advisor], Dr. Mark Lawley and Dr. Erick Moreno-Centeno of the Department of Industrial and Systems Engineering and Dr. Jianhua Huang of the Department of Statistics.

The AHA data used to build the hospital dataset for Chapters 2 and 3 was provided by Dr. Gary D. Ferrier of the Department of Economics of the University of Arkansas. Some work for the preparation of the hospital dataset was performed by Marcella Crispim-Sarmento, a former Master student of Dr. Johnson's research group in 2015. Dr. Gary D. Ferrier, introduced above, reviewed Chapters 2 and 3. Dr. Robin C. Sickles of the Department of Economics of Rice University helped to reformulate some of the concepts and when writing the application section of Chapter 2. The spline estimators used in Chapter 3 started from a discussion with Dr. Jianhua Huang who provided support when needed to write them too. My advisor Dr. Andrew Johnson contributed to all chapters of the dissertation to guide me on research ideas, concepts and reviewing the work accomplished.

All work on the dissertation is the result of the student's leadership, execution and collaboration with the above stated participants.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENTS iv

CONTRIBUTORS AND FUNDING SOURCES v

TABLE OF CONTENTS vi

LIST OF FIGURES x

LIST OF TABLES xii

1. INTRODUCTION 1

2. DIRECTION SELECTION IN STOCHASTIC DIRECTIONAL DISTANCE FUNCTIONS 5

2.1 Introduction 5
2.2 Models 8
2.2.1 Statistical Model 8
2.2.2 Production Model 10
2.2.2.1 The Deterministic Directional Distance Function (DDF) 11
2.2.2.2 The Stochastic Directional Distance Function 12
2.3 Estimation 14
2.3.1 Parametric Estimation and the DDF 14
2.3.2 The CNLS-d Estimator 15
2.4 Measuring MSE under Alternative Directions 16
2.4.1 Illustrative Example 16
2.4.1.1 Data Generation Process 16
2.4.1.2 Evaluating the Parametric Estimator's Performance 18
2.4.1.3 Additional Information Describing the Simulations 19
2.4.1.4 Results: Random Noise Directions 20
2.4.2 Radial MSE Measure 22
2.5 Monte Carlo Simulations 24
2.5.1 CNLS-d Formulation for Cost Isoquant Estimation 24
2.5.2 Experiments 25

2.5.2.1 Experiment 1 - Base case: A two output circular isoquant with uniformly distributed angle parameters and random noise direction 25
2.5.2.2 Experiment 2 - The base case with fixed noise directions 26
2.5.2.3 Experiment 3. Base case with fixed noise direction and different noise levels 27
2.5.2.4 Experiment 4: Base case with different distributions for the initial observations on the true function 28
2.6 Proposed Approach to Direction Selection 29
2.7 Cost Function Estimation of the US Hospital Sector 30
2.7.1 Description of the Data Set 30
2.7.2 Pre-Analysis of the Data Set 31
2.7.2.1 Testing the Relevance of the Regressors 31
2.7.3 Results 32
2.7.3.1 CNLS-d and Different Directions 32
2.7.3.2 Comparison with other estimators 34
2.7.3.3 Description of Functional Estimates - MPSS and Marginal Costs . 35
2.8 Conclusions 39

3. APPROXIMATION TO SHAPE CONSTRAINTS OF NON-PARAMETRIC ESTIMATORS 41

3.1 Introduction 41
3.2 Production Economics Theory and Shape Constraints 43
3.2.1 Axioms of the Cost Function 43
3.2.2 Coordinate-wise constraints versus global constraints 43
3.2.3 Afriat Inequalities and constraints validation 44
3.3 Estimators 44
3.3.1 Shape Constrained Kernel-weighted Least Squares (SCKLS) 44
3.3.1.1 Regular SCKLS 44
3.3.1.2 SCKLS with coordinate-wise constraints 45
3.3.2 Shape Constrained B-Splines Least-Squares (SCBLS) 46
3.3.2.1 B-splines introduction and univariate case 46
3.3.2.2 Expansion of B-splines to multivariate case 47
3.3.2.3 SCBLS with coordinate-wise shape constraints 47
3.3.2.4 SCBLS with local Afriat constraints 49
3.3.2.5 SCBLS with coordinate-wise constaints and local Afriat constraints 50
3.4 Monte Carlo simulations 51
3.4.1 Experiments with different data generation settings 52
3.4.1.1 Experiment 1 - Base case: A Cost Function varying the number of regressors. 52
3.4.1.2 Experiment 2 - Noise Levels variations relative to the Base case .. 55
3.4.1.3 Experiment 3 - Variation of the observation numbers relative to the Base case 56
3.4.2 Additional estimators 60

3.4.2.1 Experiment 4 - Base case analysis with partial derivatives bounds for spline 60
3.4.3 Summary and recommendations 62
3.5 Application to the cost function estimation of the US hospital sector 62
3.5.1 Presentation of the data set 63
3.5.2 Results 63
3.5.2.1 Analysis on the marginal rates of substitution 63
3.5.2.2 Marginal cost analysis 65
3.6 Conclusion 67

4. CONCLUSIONS 69

REFERENCES 71

APPENDIX A. APPENDIX OF CHAPTER 2 77

A.1 Properties of Directional Distance Functions and CNLS-d 78
A.1.1 Direction Selection in Directional Distance Functions 78
A.1.2 Details of CNLS-d 78
A.1.3 Different Directions for Different Groups in CNLS-d 79
A.1.3.1 Simulations to investigate the frequency with which multiple directions leads to violations 83
A.2 Additional Experiments 85
A.2.1 Experiments Related to Section 2.4.1 - with the Linear Estimator. 85
A.2.1.1 Measuring MSE Example, Section 2.4.1 - Noise Generated in a Common and Prespecified Direction θ_f 85
A.2.1.2 Results: Fixed Noise Direction 86
A.2.1.3 Monte Carlo Simulations - Experiments, Section 2.5.2.3 - Experiment 3. Base case with fixed noise direction and different noise levels 86
A.2.2 Experiments related to Section 2.5.2 - with CNLS-d. 86
A.2.2.1 Experiment 5: Base case with different distributions for the initial observations on the true function 87
A.2.2.2 Experiment 6: Adaptation of the Base Case to a 3-Dimensional Case 91
A.3 U.S. Hospital Dataset Application 94

APPENDIX B. APPENDIX OF CHAPTER 3 98

B.1 Formulations of SCBLS when the knots vectors are not uniform 99
B.1.1 SCBLS with coordinate-wise constraints in the univariate case in the non-uniform knots vector case 99
B.1.2 SCBLS with coordinate-wise constraints in the multivariate case in the non-uniform knots vector case 100
B.2 Additional Results for the US Hospitals Application 100
B.2.1 Estimators performance for a cost function with four outputs. 100

B.2.2 Fitting performance with aggregated outputs 101
B.2.3 Results obtained with SCKLS 102

LIST OF FIGURES

FIGURE Page

2.1 SDDF in mean zero case 13

2.2 Algorithm 1: Linear function data generation process with random noise directions . 17

2.3 Linear Case with Random Noise Direction (left), Linear Case with Fixed Noise Direction (right) 18

2.4 MSE calculated relative to the True Function in the MSE direction $\pi/4$ (left), MSE calculated using a testing data set in the MSE direction $\pi/4$ (right) 19

2.5 A Radial MSE Measure on a Cost Function with Two Outputs 23

2.6 US Hospital Cost Function Estimates for Three Directions 33

2.7 Most Productive Scale Size (in red) on the estimated function by “CNLS-d (Med)", CNLS-d using the median approach for the direction, for different ratios of Major Therapeutic Procedures over Minor Therapeutic Procedures 37

2.8 Marginal Cost of the Minor Therapeutic procedures (left) and Marginal Cost of the Major Therapeutic procedures (right) (“CNLS-d (Med)" corresponds to CNLS-d using the median approach for the direction and “CNLS-d (Eq)” corresponds to CNLS-d using the direction with equal components in all netputs 38

3.1 MSE in the base case with 100 observations and for a noise level with standard deviation $\sigma = 0.1$ on 50 simulations 54

3.2 Percentage of Violations of the Afriat Inequalities concerning convexity on the set of testing points 54

3.3 MSE measured relative to the true function for a set of 100 testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is 100 measured over 50 simulations. 56

3.4 Percentage of violations of Afriat convexity for a set of 100 testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is 100 measured over 50 simulations. 57

3.5 MSE measured relative to the true function for a set of $\{100,\ 200,\ 500\}$ testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is $\{100,\ 200,\ 500\}$ measured over 50 simulations. 58

3.6 Percentage of violations of Afriat convexity for a set of $\{100,\ 200,\ 500\}$ testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is $\{100,\ 200,\ 500\}$ measured over 50 simulations. 59

3.7 MSE measured relative to the true function for a set of 100 testing points with the noise standard deviation $\sigma = 0.1$ and the number of observations is 100 measured over 50 simulations adding results from estimator with slope bounds..... 61

3.8 Percentage of violations of Afriat convexity for a set of 100 testing points with the noise standard deviation $\sigma = 0.1$ and the number of observations is 100 measured over 50 simulations adding results from estimator with slope bounds..... 61

3.9 MRS between Major Therapeutic procedures and Minor Therapeutic procedures obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values. 65

3.10 MC of the Major Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values..... 66

3.11 MC of the Minor Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values..... 66

A.1 Linear function data generation process with fixed noise direction 86

A.2 (a) On the left, histogram of the angles θ_i and its median obtained for each distribution when running a simulation with 100 observations. (b) On the right, the corresponding observations and the median of the angles θ_i for each distribution for the same simulation as the histogram (a)..... 89

A.3 Experiment 6: 3-dimensional isoquant case: Representation of the directions tested and the values of averaged radial MSE, the size of the markers having a positive affine relation to the values. The color is an other indicator as the more red the higher the averaged radial MSE is and the more yellow for lower values..... 92

B.1 MRS between Major Therapeutic procedures and Minor Therapeutic procedures obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values..... 102

B.2 MC of the Major Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values. 103

B.3 MC of the Minor Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values. 103

LIST OF TABLES

TABLE Page

2.1 Average MSE over 100 simulations for the Linear Estimator compared to the true function with a DGP using random noise directions 20

2.2 Average MSE over 100 simulations for the Linear Estimator compared to an out-of-sample testing set with a DGP using random noise directions 21

2.3 Experiment 1: Values of the radial MSE relative to the true function. The angle used in CNLS-d estimator varies and the noise direction is randomly selected. In the DGP, the standard deviation of the noise distribution, λ, is 0.1. 26

2.4 Experiment 2: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d estimator direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.1. 27

2.5 Experiment 3–Less Noise: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.05. 27

2.6 Experiment 4: Values of radial MSE relative to the true function varying the CNLS-d direction and the mean of the normal distribution used in the DGP. 28

2.7 Summary Statistics of the Hospital Data Set 31

2.8 Results of the radial MSE values for different directions by year 33

2.9 US Hospital K-fold Average MSE in Cost to the Cost Function Estimates for the Three Functional Specifications by Year 34

2.10 Most Productive Scale Size measured in cost ($\$M$) conditional on Minor Therapeutic procedures (MinTher) and Major Therapeutic procedures (MajTher), Minor Diagnostic procedures (MinDiag) and Major Diagnostic procedures (MajDiag) held constant at the 50th percentile 36

2.11 Marginal Cost of Minor Therapeutic Procedures 37

2.12 Marginal Cost of Major Therapeutic Procedures 38

3.1 The four estimators considered to establish a comparison of the different approximations to the shape constraints. The estimators in italics are created for this analysis. 43

3.2 Optimal estimators in a case of non-extreme noise levels based on our simulation results depending of data sets characteristics 63

3.3 Data Summary 64

A.1 Average MSE over 100 simulations for the Linear Estimator compared to the true function with a DGP using random noise directions 87

A.2 Average MSE over 100 simulations for the Linear Estimator compared to an out-of-sample testing set with a DGP using fixed noise directions 88

A.3 Experiment 3–More Noise: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.2. 89

A.4 Experiment 5: Values of radial MSE relative to the true function varying the CNLS-d direction and type of direction used for the DGP. 90

A.5 Experiment 6: Values of radial MSE relative to the true function varying the CNLS-d direction in a 3- dimensional case 93

A.6 Most Productive Scale Size (c) 95

A.7 Marginal Cost of Minor Therapeutic Procedures 96

A.8 Marginal Cost of Major Therapeutic Procedures 97

B.1 Averaged K-fold MSE to the observations with its standard deviation (underneath in italics) on two years of the hospital data obtained for different estimators. 101

B.2 Averaged K-fold MSE to the observations with its standard deviation (underneath in italics) on two years of the hospital data obtained for different estimators. 102

1. INTRODUCTION

As datasets grow larger and computation power increases, data analytics is expanding as well. This dissertation contributes to researchers understanding of newly developed methods, shape-constrained non-parametric, that give more reliable estimations and improve analytic quality. Improvements in performance assessment methods will benefit both policy-makers and decision-makers in manufacturing, finance, education, etc.

A few prominent applications illustrate the importance of performance measurements. For example, the electricity distribution market in several European countries is privatized, firms have local monopoly power to build and operate the last miles of transmission lines, and prices are controlled by a regulator. Since 2012, the Finnish regulator has used a framework for rewarding efficient operators and increasing the network efficiency based on Stochastic Nonparametric Envelopment of Data (StoNED). For an introduction to the framework see Kuosmanen and Kortelainen (2012) and Kuosmanen (2012), and for details about StoNED see Kuosmanen et al. (2015) and Kuosmanen and Johnson (2017). Another popular regulatory application is carbon permits that several countries have implemented to reduce CO_2 emissions. In China, companies in specific industries known to release vast quantities of emissions CO_2 receive permits that allow them to release a certain volume of CO_2 emissions. If they need to emit more CO_2, they can purchase additional permits from other companies or from the government. This creates a regulated carbon trading market, see for example Zhao et al. (2016). Production economics is used to ensure the industry transitioning into a net zero carbon footprint by estimating the lost output if companies choose to abate by reducing output, or by estimating an appropriate price for additional carbon permits purchased from the government. Another set of applications concerns the evaluation of management practices. The “World Management Survey" described in Bloom and Van Reenen (2007) and Bloom and Van Reenen (2010) relates management practices to productivity. Bloom and Van Reenen (2010) find a significant correlation between good management (as measured by their survey) and higher productivity levels for manufacturing facilities, hospitals, and schools

around the world.

The US healthcare industry could gain from more productivity and efficiency analyses; in 2016, health expenses per capita in the US were 31% higher than in Switzerland, which had the second-largest national per capita value. The US hospital industry is the focus of the application subsections of both main chapters of this dissertation. Using the National Inpatient Sample (NIS) from the Healthcare Cost and Utilization Project (HCUP) and data from the American Hospital Association (AHA) Annual Survey, the cost function for US hospitals is estimated for multiple years. From the cost function some of the metrics determined are for instance the most productive scale sizes (MPSS) of the hospitals. The MPSS are measured for different output ratios and all values estimated are small. Thus one of the results obtained is that the small hospitals, that also represent most of the operating hospitals, appear to be more productive than large hospitals. Based on this analysis and other completing analyses, decision-makers can design better hospital networks to improve the cost effectiveness of the system.

As described above cost functions are used to determine valuable insights. A cost function represents the levels of inputs, measured by their cost, to achieve different combinations of outputs and thus characterizes a production technology. For example, the outputs for a hospital are the number of procedures for different categories, and the cost is the hospitals total expenses. In the dissertation we focus our research on the estimation of the conditional mean of the production technology. We refer the reader to Kuosmanen and Johnson (2017), that integrates the method of Hall and Simar (2002) to estimate the frontier from the conditional mean to determine the efficiency levels.

To estimate production technologies many researchers use deterministic models like Data Envelopment Analysis (DEA). In this dissertation we prefer the stochastic model that is more realistic. In particular in Chapter 2, the statistical model allows the possibility of noise in all observations and for measurement errors in all components. Using a stochastic model also raises questions about endogeneity, that corresponds to the error term being correlated to the regressors of the model. In Chapter 2, we use the Stochastic Directional Distance Functions for estimating multi-output

cost functions and potentially improving the estimations in some endogeneous cases. The chapter shows that the choice of the direction used for the estimator using the Directional Distance Function matters and it provides recommendations for the user to select the direction of the estimator using stochastic directional distance functions.

An other aspect of production technologies estimation is that many researchers still use parametric models such as translog or Cobb-Douglas functional forms. However, parametric estimators seem to rely on strict assumptions and can suffer from misspecification in most applications because the correct functional form is typically unknown and cannot be determined by the observable characteristics of production. On the other hand, fully non-parametric models like local linear kernel regressions tend to overfit the observed sample and provide limited out-of-sample information or structural information. Imposing shape constraints allows for a compromise between parametric and nonparametric models. For instance, cost functions are commonly assumed to be non-decreasing and convex for any increase in output levels, i.e., it takes additional effort for a hospital to increase its number of procedures. Shape-constrained non-parametric estimators such as CNLS (Kuosmanen (2008)), SCKLS (Yagi et al. (2018)), MBCR (Hannah and Dunson (2013)) and CWB (Du et al. (2013)) use different approximations of the shape constraints or functions or both.

In Chapter 3, the approximation piece-wise linear versus smooth (at least of class $\mathbf{C}^2$) estimated functions and the approximation of coordinate-wise shape constraints versus global constraints are analyzed. Hence, an estimator like SCKLS is piece-wise linear but has global shape constraints while an estimator like those in Pya and Wood (2015) is smooth but only satisfies coordinate-wise constraints. To evaluate the effect of each approximation, two estimators are created: SCKLS with coordinate-wise constraints, which is a piece-wise linear estimator with coordinate-wise constraints, and a spline estimator derived from Pya and Wood (2015) with global shape constraints imposed on a set of control points, which is smooth with some global constraints. The effects of each approximation can be evaluated separately. Two criteria are used to evaluate each estimator: goodness of fit and non-violations of the shape constraints. Based on the results of Monte Carlo

experiments, Chapter 3 ends by providing guidelines to help users select the best estimator among the ones considered.

The remainder of the dissertation is structured as follows. Chapter 2 addresses the issue of direction selection in stochastic directional distance functions, questions related to the problem, and provides recommendations. Chapter 3 evaluates several approximations to shape constraints of non-parametric estimators, and suggests guidelines to help users' estimator selection. Chapters 2 and 3 also give an example of an application for the US hospitals industry. Chapter 4 concludes.

2. DIRECTION SELECTION IN STOCHASTIC DIRECTIONAL DISTANCE FUNCTIONS *

2.1 Introduction

The focus of this chapter is direction selection in stochastic directional distance functions (SDDF).[2] While the DDF is typically used to measure efficiency, in this chapter we use a nonparametric shape constrained SDDF to model the conditional mean behavior of production. The stochastic distance function (SDF) was introduced by Lovell et al. (1994) and was used in a series of early empirical studies by Coelli and Perelman (1999, 2000) and Sickles et al. (2002). The parameters of a parametric distance function are point identified; however, if the direction in the DDF is not specified, then the parameters of a parametric DDF are set identified.[3] A set of axiomatic properties related to production and cost functions, such as monotonicity and convexity in the case of a cost function, are well established in the production literature (Shephard (1970); Chambers (1988)). Although the stochastic distance function literature acknowledges the axiomatic properties necessary for duality, it does not impose them globally. Instead, authors typically impose them only on a particular point in the data (e.g., Atkinson et al. (2003)). Recognizing these issues, we provide an axiomatic nonparametric estimator of the SDDF and a method to restrict the pool of the directions to choose from for the SDDF, thereby reducing the size of the set identified parameter set. Most empirical studies that use establishment or hospital level data to estimate production or cost functions either assume a specific parametric form or ignore noise, or both ((Hollingsworth, 2003)). In contrast, we use an axiomatic nonparametric SDDF estimator and the proposed method to determine a set of acceptable directions to estimate a cost function that maintains global axiomatic properties for the US hospital industry. Furthermore, we demonstrate the importance of

*Reprinted with permission from "Direction Selection in Stochastic Directional Distance Functions" by Layer, K., Johnson, A. L., Sickles, R. C., & Ferrier, G. D., 2020. *European Journal of Operational Research*, 280(1), 351-364, Copyright 2020 by Elsevier

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

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

global axiomatic properties for the estimation of most productive scale size and marginal costs.

A few papers have attempted to implement the directional distance function in a stochastic setting (see, for example, Färe et al. (2005), Färe et al. (2010), and Färe and Vardanyan (2016)). The latter two papers discuss the challenges of selecting a parametric functional form that does not violate the axioms typically assumed in production economics. Based on their observations, Färe and Vardanyan (2016) use a quadratic functional specification.[4] Yet several papers show a loss of flexibility in parametric functional forms, such as the translog or the quadratic functional form, when shape constraints are imposed (e.g., Diewert and Wales (1987)). Also important to implementation, the selection of the direction vector in the SDDF has been discussed in Färe et al. (2017) and Atkinson and Tsionas (2016), among others. These papers focus on selecting the direction corresponding to a particular interpretation of the inefficiency measure, based on the distance to the economically efficient point. In contrast, we consider Kuosmanen and Johnson (2017)'s multi-step efficiency analysis and focus on the first step, estimating a conditional mean function. Our goal is to select the direction that best recovers the underlying technology while acknowledging that the data is likely to contain noise in potentially all variables.[5]

To model multi-product production, Kuosmanen and Johnson (2017) have proposed the use of axiomatic nonparametric methods to estimate the SDDF which they name Directional Convex Nonparametric Least Squares (CNLS-d), a type of sieve estimator. Their methods have the benefits of relaxing standard functional form assumptions for production, cost, or distance functions, but also improve the interpretability and finite sample efficiency over nonparametric methods such as kernel regression (Yagi et al. (2018)). A variety of models can be interpreted as special cases of Kuosmanen and Johnson (2017), among these are a set of models that specify the direction (e.g., Johnson and Kuosmanen (2011); Kuosmanen and Kortelainen (2012)). All CNLS models are sieve

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

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

estimators and fall into the category of partially identified or set identified estimators discussed in Manski (2003) and Tamer (2010). The guidance our paper provides in selecting a direction will reduce the size of the set identified for CNLS-d and other DDF estimators with flexible direction specifications.

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

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

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

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

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

2.2 Models

2.2.1 Statistical Model

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

$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{y}}_i \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}^x_i \\ \boldsymbol{\epsilon}^y_i \end{pmatrix}. \tag{2.1}$$

Equation (2.1) is only identified when multiple measurements exist for the same vector of regressors or when a subsample of observations exists in which the regressors are measured exactly (Carroll et al. (2006)). Carroll et al. (2006) discussed a standard regression setting, not a multi-input/multi-output production process. Thus, repeated measurement requires all but one of the

netputs to be identical across at least two observations.[6] Neither of of these conditions is likely to hold for typical production data sets; therefore, we develop an alternative approach to identification.

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

$$\begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{x}}_i \\ \tilde{\boldsymbol{y}}_i \end{pmatrix} + e_i \begin{pmatrix} \boldsymbol{g}_i^x \\ \boldsymbol{g}_i^y \end{pmatrix}. \tag{2.2}$$

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

$$\begin{pmatrix} \boldsymbol{\epsilon}_i^x \\ \boldsymbol{\epsilon}_i^y \end{pmatrix} = e_i \begin{pmatrix} \boldsymbol{g}_i^x \\ \boldsymbol{g}_i^y \end{pmatrix}. \tag{2.3}$$

We define the following normalization:

$$e_i = \sqrt{\sum_{j=1}^{d} (\epsilon_{ij}^x)^2 + \sum_{j=1}^{Q} (\epsilon_{ij}^y)^2}, \tag{2.4}$$

which implies:

$$\sqrt{\sum_{j=1}^{d} (g_{ij}^x)^2 + \sum_{j=1}^{Q} (g_{ij}^y)^2} = 1. \tag{2.5}$$

We refer to $(\boldsymbol{g}_i^x, \boldsymbol{g}_i^y)$ as the *true* noise direction and in the most general case we allow the direction to be observation specific.[7] The estimation methods to consider noise in potentially all inputs will depend on our assumptions about the production technology, which are discussed in the following subsection.

[6] Here we use the term *netputs* to describe the union of the input and output vectors.

[7] When the noise direction is observation specific and random, all inputs and outputs potentially contain noise and therefore are endogeneous variables. If some components of the $(\boldsymbol{g}^x, \boldsymbol{g}^y)$ vector are zero, this implies the associated variables are exogeneous and measured with certainty. See Kuosmanen and Johnson (2017) for more details.

2.2.2 Production Model

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

$$T = \left\{ (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathbb{R}_{+}^{d+Q} \mid \tilde{\boldsymbol{x}} \text{ can produce } \tilde{\boldsymbol{y}} \right\}. \tag{2.6}$$

Following Shephard (1970), we adopt the following standard assumptions to assure that T represents a production technology:

1. T is closed;

2. T is convex;

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

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

Developing methods to estimate characteristics of the production technology while imposing these standard axioms was a popular and fruitful topic from the early 1950's until the early 1980's, generating such classic papers as Koopmans (1951), Shephard (1953, 1970), Afriat (1972), Charnes et al. (1978),[8] and Varian (1984). Unfortunately, these methods are deterministic in the sense that they rely on a strong assumption that the data do not contain any measurement errors, omitted variables, or other sources of random noise. Furthermore, for some research communities linear programs were seen as harder to implement than parametric regression which could be calculated via normal equations. Thus, most econometricians and applied economists have chosen to use parametric models, sacrificing flexibility for ease of estimation and the inclusion of noise in the model.

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

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

2.2.2.1 The Deterministic Directional Distance Function (DDF)

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

$$\overrightarrow{D}_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) = \max\left\{\delta \in \mathbb{R} : \left(\tilde{\boldsymbol{x}} - \delta\, \boldsymbol{g}^x,\ \tilde{\boldsymbol{y}} + \delta\, \boldsymbol{g}^y\right) \in T\right\}, \tag{2.7}$$

where $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ are the observed input and output vectors, such that $\tilde{\boldsymbol{x}} \in \mathbb{R}^d_+$ and $\tilde{\boldsymbol{y}} \in \mathbb{R}^Q_+$ are assumed to be observed without noise and fully describe the resources used in production and the goods or services generated from production. $\boldsymbol{g}^x \in \mathbb{R}^d_+$ is the direction vector in the input space, $\boldsymbol{g}^y \in \mathbb{R}^Q_+$ is the direction vector in the output space, and $(\boldsymbol{g}^x, \boldsymbol{g}^y) \in \mathbb{R}^{d+Q}_+$ defines the direction from the point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ in which the distance function is measured.[9] δ is commonly interpreted as a measure of inefficiency by quantifying the number of bundles of size $(\boldsymbol{g}^x, \boldsymbol{g}^y)$ needed to move the observed point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ to the boundary of the technology in a deterministic setting.

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

$$\overrightarrow{D}_T\left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y\right) \geq 0, \text{if and only if } \left(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\right) \in T. \tag{2.8}$$

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

If T satisfies the assumptions stated in Section 2.2.2, then the directional distance function $\overrightarrow{D}_T : \mathbb{R}_+^d \times \mathbb{R}_+^Q \times \mathbb{R}_+^d \times \mathbb{R}_+^Q \to \mathbb{R}_+$ has the following properties (see Chambers et al. (1998)):

(a) $\overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y)$ is upper semicontinuous in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{y}}$ (jointly);

(b) $\overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \lambda\, \boldsymbol{g}^x, \lambda\, \boldsymbol{g}^y) = (1/\lambda)\, \overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y), \lambda > 0$;

(c) $\tilde{\boldsymbol{y}}' \geq \tilde{\boldsymbol{y}} \Rightarrow \overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}'; \boldsymbol{g}^x, \boldsymbol{g}^y) \leq \overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y)$;

(d) $\tilde{\boldsymbol{x}}' \geq \tilde{\boldsymbol{x}} \Rightarrow \overrightarrow{D}_T(\tilde{\boldsymbol{x}}', \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y) \geq \overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y)$;

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

An additional property of the DDF is the translation invariance:

(f) $\overrightarrow{D}_T(\tilde{\boldsymbol{x}} - \alpha \boldsymbol{g}^x, \tilde{\boldsymbol{y}} + \alpha \boldsymbol{g}^y; \boldsymbol{g}^x, \boldsymbol{g}^y) = \overrightarrow{D}_T(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}; \boldsymbol{g}^x, \boldsymbol{g}^y) - \alpha$.

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

2.2.2.2 The Stochastic Directional Distance Function

The properties of the deterministic DDF also apply for the stochastic DDF (Färe et al. (2017)). Here we focus on estimating a stochastic DDF considering a residual which is mean zero.[10] This is represented in Figure 2.1.

Using the statistical model in Section 2.2.1 and the functional representation of technology in Section 2.2.2, we restate Proposition 2 in Kuosmanen and Johnson (2017) as:

Proposition 1. *If the observed data are generated according to the statistical model described in Section 2.2.1, then the value of the DDF in the observed data point* $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ *is equal to the*

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

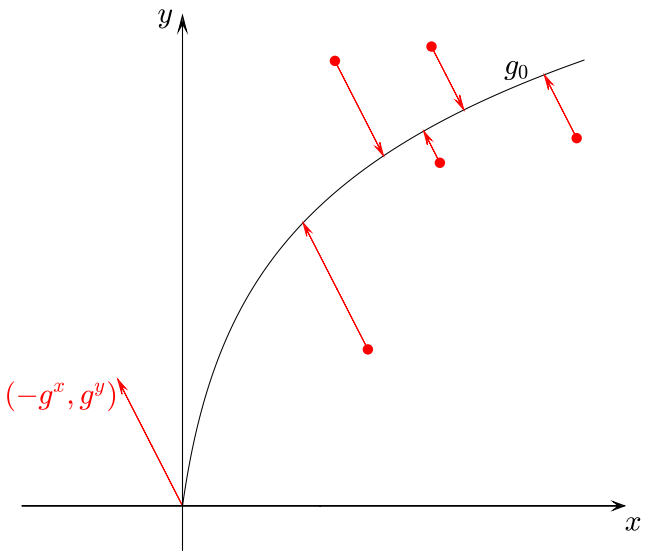

Figure 2.1: SDDF in mean zero case

realization of the random variable ϵ_i with mean zero; specifically

$$\overrightarrow{D}_T(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{g}^x, \boldsymbol{g}^y) = \epsilon_i \; \forall i.$$

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

Under the assumption of constant returns to scale, normalizing by one of the netputs causes the noise terms to cancel for the regressors, thus eliminating the issue of endogeneity (e.g., Coelli (2000), Kuosmanen and Johnson (2017)). However, since we relax the constant returns to scale assumption, endogeneity can still be an issue.[11]

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

Färe et al. (2017), among others, have recognized that the selection of the direction vector affects the parameter estimates of the production function. In Appendix A.1.1, for the linear parametric DDF defined below, we prove that alternative directions lead to distinct parameter estimates.

2.3 Estimation

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

2.3.1 Parametric Estimation and the DDF

Consider data composed of n observations where the inputs are defined by $\boldsymbol{x_i},\ i = 1, ..., n$ and the outputs by $\boldsymbol{y_i},\ i = 1, ..., n$. The estimator minimizes the squared residuals for a DDF with an arbitrary prespecified direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$. For a linear production function, we formulate the estimator as:

$$\min_{\alpha,\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{2.9}$$

$$\text{s.t.} \quad \boldsymbol{\gamma}'\, \boldsymbol{y_i} = \alpha + \boldsymbol{\beta}'\, \boldsymbol{x_i} - \epsilon_i, \text{ for } i = 1, \ldots, n \tag{2.9a}$$

$$\boldsymbol{\beta}'\, \boldsymbol{g}^x + \boldsymbol{\gamma}'\, \boldsymbol{g}^y = 1, \tag{2.9b}$$

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

Equation (2.9b) enforces the translation property described in Chambers et al. (1998); i.e., scaling the netput vector by δ in the direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$ causes the distance function to decrease by δ. The combination of Equation (2.9a) and Equation (2.9b) ensures that the residual is computed along the direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$. Intuitively this is because the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are rescaled proportionally to the direction $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$ in Equation (2.9b). For a formal proof, see Kuosmanen and Johnson (2017), Proposition 2.

below in Section 2.4.1.

2.3.2 The CNLS-d Estimator

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

Using the same input-output data defined in Section 2.2.1, the CNLS-d estimator is given by:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{2.10}$$

$$\text{s.t.} \quad \boldsymbol{\gamma_i'}\,\boldsymbol{y_i} = \alpha_i + \boldsymbol{\beta_i'}\,\boldsymbol{x_i} - \epsilon_i, \text{ for } i = 1, \ldots, n \tag{2.10a}$$

$$\alpha_i + \boldsymbol{\beta_i'}\,\boldsymbol{x_i} - \boldsymbol{\gamma_i'}\,\boldsymbol{y_i} \leq \alpha_j + \boldsymbol{\beta_j'}\,\boldsymbol{x_i} - \boldsymbol{\gamma_j'}\,\boldsymbol{y_i}, \text{ for } i, j = 1, \ldots, n,\ i \neq j \tag{2.10b}$$

$$\boldsymbol{\beta_i} \geq 0, \text{ for } i = 1, \ldots, n \tag{2.10c}$$

$$\boldsymbol{\beta_i'}\,\boldsymbol{g^x} + \boldsymbol{\gamma_i'}\,\boldsymbol{g^y} = 1, \text{ for } i = 1, \ldots, n \tag{2.10d}$$

$$\boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1, \ldots, n, \tag{2.10e}$$

where $\alpha_i,\ i = 1, ..., n$ is the vector of the intercept terms, $\boldsymbol{\beta_i},\ i = 1, .., n$ and $\boldsymbol{\gamma_i},\ i = 1, .., n$ are the matrices of the marginal effects of the inputs and the outputs, respectively, and $\epsilon_i,\ i = 1, ..., n$ is the vector of the residuals (Kuosmanen and Johnson, 2017).

Equation (2.10a) is similar to (2.9a) with the notable difference that $(\alpha_i, \boldsymbol{\beta_i}, \boldsymbol{\gamma_i})$ are indexed by i indicating each observation has their own hyperplane defined by the triplet $(\alpha_i, \boldsymbol{\beta_i}, \boldsymbol{\gamma_i})$. Equation (2.10b), which corresponds to the Afriat inequalities, imposes concavity. Given Equation (2.10b),

Equation (2.10c) imposes the monotonicity of the estimated frontier relative to the inputs. Equation (2.10d) enforces the translation property described in Chambers et al. (1998) and has the same interpretation as Equation (2.9b). Similar to Equation (2.10c), the combination of Equation (2.10b) and Equation (2.10e) imposes the monotonicity of the DDF relative to the outputs. In Equation (2.10), we specify the CNLS-d estimator with a single common direction, $(-\boldsymbol{g}^x, \boldsymbol{g}^y)$.[12]

2.4 Measuring MSE under Alternative Directions

2.4.1 Illustrative Example

2.4.1.1 Data Generation Process

For our illustrative example, we use a simple linear cost function and a directional distance linear parametric estimator. We consider two noise generation processes: a random noise direction and a fixed noise direction. Here we discuss the random noise direction case, but direct the reader to Appendix A.2 for a discussion of the fixed noise direction case.

For our example we consider a single output cost function where the observations $(y_i, c_i), i = 1, \ldots, n$, are created by the Data Generation Process (DGP) outlined in Algorithm 1:

[12] Alternatively, some researchers may be interested in using observation specific directions or perhaps group specific directions (Daraio and Simar (2016)). In Appendix A.1.3, we derive the conditions under which multiple directions can be used in CNLS-d while still maintaining the axiomatic property of global convexity of the production technology. Consider two groups each with their own direction used in the directional distance function. Essentially, the convexity constraint holds as long as the noise is orthogonal to the difference of the two directions used in the estimation. A simple example of this situation is all the noise being in one dimension and the difference between the two directions for this dimension is zero. However, this condition is restrictive when noise is potentially present in all variables. Thus, specifying multiple directions in CNLS-d while maintaining the axiomatic properties of the estimator, specifically, the convexity of the production possibility set, is still an open research question.

Algorithm 1

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

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

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

   (a) ϵ_0 is calculated as:

   $$\epsilon_0 = \frac{1}{2}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2} + \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{c}_i - \bar{c}\right)^2}\right], \tag{2.11}$$

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

   (b) The scalar length of the noise is rescaled by the vector, $v_{q\epsilon_i}$, in each dimension. These scaling factors are calculated as $v_{q\epsilon_i} = \frac{v^*_{q\epsilon_i}}{||\boldsymbol{v}^*_{\boldsymbol{\epsilon_i}}||_2}, q = \{1, 2\}$ where $v^*_{q\epsilon_i}$ are drawn from a continuous uniform distribution $U[-1, 1]$.

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

4. The observations with noise are obtained by appending the noise terms to the generated data:

   $$\begin{pmatrix} y_i \\ c_i \end{pmatrix} = \begin{pmatrix} \tilde{y}_i \\ \tilde{c}_i \end{pmatrix} + \begin{pmatrix} \epsilon_{y_i} \\ \epsilon_{c_i} \end{pmatrix}, i = 1, \ldots, n. \tag{2.12}$$

Figure 2.2: Algorithm 1: Linear function data generation process with random noise directions

Figure 2.3 illustrates the results for two cases of the data generating process; in the first case the

direction of the noise is random, while in the second case the direction of the noise is fixed.

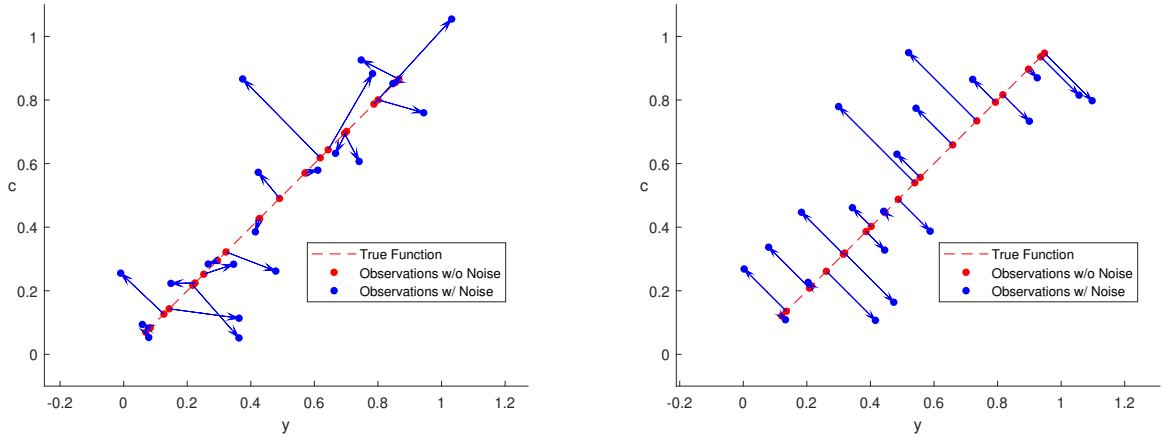

Figure 2.3: Linear Case with Random Noise Direction (left), Linear Case with Fixed Noise Direction (right)

2.4.1.2 Evaluating the Parametric Estimator's Performance

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

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

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left((\hat{y}_{ts_i} - y_{ts_i})^2 + (\hat{c}_{ts_i} - c_{ts_i})^2 \right). \tag{2.13}$$

To compare the true function to the estimated function, we use the Linear Function Data Gener-

ation Process, Algorithm 1, steps 1 and 2, to construct our testing data set $(y_{ts_i}, c_{ts_i}), i = 1, \ldots, n$. To evaluate the estimated function without knowing the true function the testing set is built using the full Linear Function Data Generation Process.

Figure 2.4 show the MSE computations.

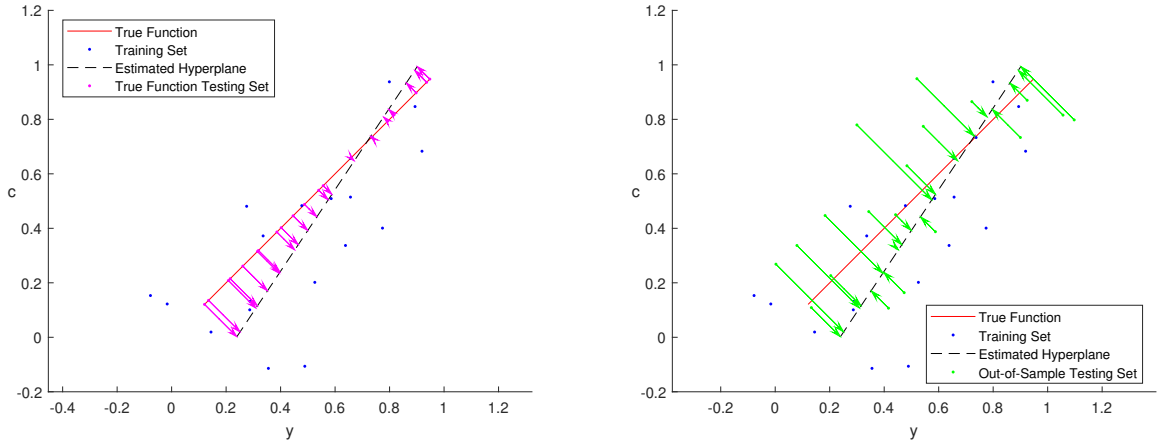

Figure 2.4: MSE calculated relative to the True Function in the MSE direction $\pi/4$ (left), MSE calculated using a testing data set in the MSE direction $\pi/4$ (right)

2.4.1.3 Additional Information Describing the Simulations

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

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

2.4.1.4 Results: Random Noise Directions

Table 2.1 and Table 2.2 show results corresponding to the two performance criteria introduced above and shown in Figure 2.4, the MSE relative to the true function and the MSE relative to a testing data set, respectively. Table 2.1 shows that the direction corresponding to the angle $\pi/4$, $(g^y = 0.707, g^c = 0.707)$, produces the smallest values of MSE (shown in bold in the table) regardless of the direction used for the MSE computation. However, the estimator's quality diminishes if we select the extreme directions corresponding to the angles 0 and $\pi/2$. Table 2.2 reports performance via a testing set, the direction corresponding to the smallest MSE value (shown in bold) is always the one matching the direction used in the MSE computation. In applications, using a testing set is necessary because the true function is unknown. Table 2.2 shows the benefits of matching the direction of MSE evaluation direction outweigh the benefits of selecting a direction based on the properties of the function being estimated.

Table 2.1: Average MSE over 100 simulations for the Linear Estimator compared to the true function with a DGP using random noise directions

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

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

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

Table 2.2: Average MSE over 100 simulations for the Linear Estimator compared to an out-of-sample testing set with a DGP using random noise directions

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

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

be used in the MSE evaluation would produce a relatively low MSE compared to other directions. However, when the functional estimate is compared to the true function, the MSE values are around ten times smaller than the out-of-sample testing case. In out-of-sample testing the presence of noise in the observations causes a deviation regardless of the quality of the estimator or the number of observations. The DDF direction corresponding to the smallest MSE is the direction orthogonal to the true function (i.e., $\pi/4$ for our DGP). This direction provides the shortest distance from the observations to the true function. We conclude that, in this experiment, it is preferable to select a direction orthogonal to the true function (see Section 2.5 for further experiments).

From the fixed noise direction experiments (see Appendix A.2.1), we observe that using a direction for the estimator that matches the direction used for the noise generation significantly reduces the MSE values compared to the true function. From this, we infer that when endogeneity is severe, using a direction that matches the characteristics of this endogeneity significantly improves the fit of the estimator; i.e., the MSE is 50% smaller for the matching direction than for the second best direction in 70% of the cases (see Section 2.5 for the details).

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

2.4.2 Radial MSE Measure

MSE is typically measured by the average sum of squared errors in the dimension of a single variable, such as cost or output. As explained in Section 2.4.1, when we compare out-of-sample performance, we find that the best direction to use in estimating a SDDF is the direction used for MSE evaluation regardless of the direction of noise in the DGP or any other characteristics of the DGP. To avoid this relationship between the direction of estimation and the direction of evaluation, we propose a **radial MSE measure**.

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

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

The normalized observations are:

$$\breve{y}_{ij} = \frac{y_{ij} - \min_k y_{kj}}{\max_k y_{kj} - \min_k y_{kj}},\ j = 1, \ldots, Q,\ i, k = 1, \ldots, n, \tag{2.14}$$

$$\breve{c}_i = \frac{y_i - \min_k c_k}{\max_k c_k - \min_k c_k},\ i, k = 1, \ldots, n. \tag{2.15}$$

Our radial MSE measure is the distance from the testing set observation to the estimated function measured along a ray from the testing set observations to the **center** C. Having normalized the data, the center for the radial measure is $C = [\breve{y}_1, , \breve{y}_Q, \breve{c}] = \left[\overbrace{0, \ldots, 0}^{Q}, 1\right]$.

The radial MSE measure is the average of the distance from each testing set observation to the estimated function measured radially. Figure 2.5 illustrates this measure. For a convex function, a radial measure reduces the bias in the measure for extreme values in the domain.

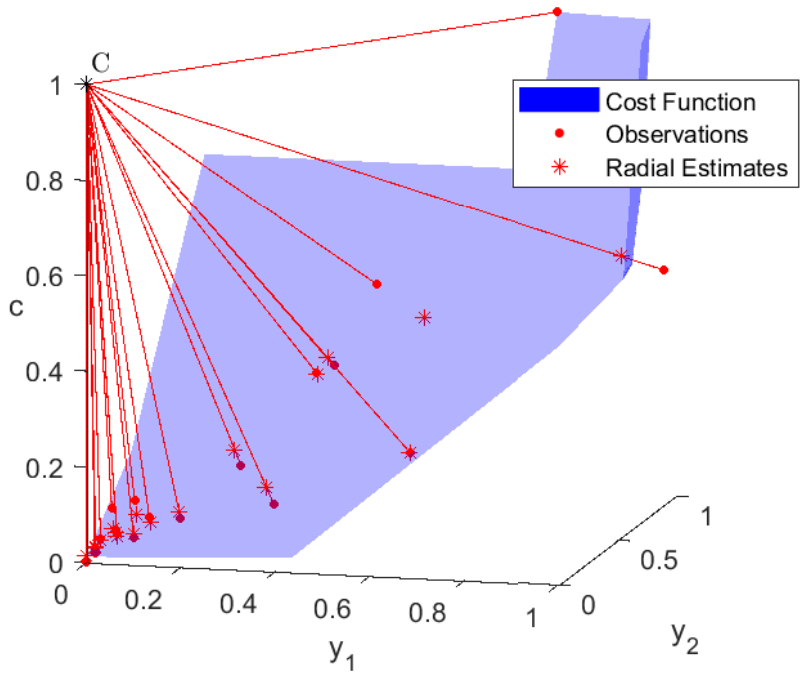

Figure 2.5: A Radial MSE Measure on a Cost Function with Two Outputs

2.5 Monte Carlo Simulations

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

2.5.1 CNLS-d Formulation for Cost Isoquant Estimation

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

$$\min_{\boldsymbol{\gamma},\boldsymbol{\epsilon}} \sum_{i=1}^{n} \epsilon_i^2 \tag{2.16}$$

$$\text{s.t.} \quad -\epsilon_j + \epsilon_i - \boldsymbol{\gamma_i'}\,(\boldsymbol{y_i} - \boldsymbol{y_j}) \leq 0, \text{ for } i, j = 1, \ldots, n,\ i \neq j \tag{2.16a}$$

$$\boldsymbol{\gamma_i'}\,\boldsymbol{g}^y = 1, \text{ for } i = 1, \ldots, n \tag{2.16b}$$

$$\boldsymbol{\gamma_i} \geq 0, \text{ for } i = 1, \ldots, n. \tag{2.16c}$$

Note all observations, $\boldsymbol{y}_i$, have a common cost level. This allows us to focus on a 2-dimensional estimation problem. For results related to 3-dimensional estimation problems see Appendix A.2.2, Experiment 6.

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

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{y_i} - \epsilon_i\,\boldsymbol{g}^y, \text{ for } i = 1, \ldots, n \tag{2.17}$$

$$\alpha_i = \boldsymbol{\gamma_i'}\,\boldsymbol{y_i} + \epsilon_i, \text{ for } i = 1, \ldots, n. \tag{2.18}$$

2.5.2 Experiments

We conducted several experiments to investigate the optimal direction for the DDF estimator. Four experiments' results are shown in the main text of the chapter with two additional experiments described in the appendix.

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

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

$$y_{qi} = \tilde{y}_{qi} + \epsilon_{qi},\ q = 1, \ldots, Q,\ i = 1, \ldots, n, \tag{2.19}$$

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

$$\tilde{y}_{1i} = \cos(\theta_i),\ i = 1, \ldots, n \tag{2.20}$$

$$\tilde{y}_{2i} = \sin(\theta_i),\ i = 1, \ldots, n, \tag{2.21}$$

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

$$\epsilon_{1i} = l\ \cos(\theta_{\epsilon_i}),\ i = 1, \ldots, n \tag{2.22}$$

$$\epsilon_{2i} = l\ \sin(\theta_{\epsilon_i}),\ i = 1, \ldots, n, \tag{2.23}$$

where the length l is drawn from the normal distribution $N(0, \lambda)$, the angle θ_{ϵ_i} is observation specific and characterizes the noise direction for each observation, and θ_{ϵ_i} is drawn from a continuous

uniform distribution $U\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. The values considered for the directions in CNLS-d estimator are $\theta_{\textit{CNLS-d}} \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}\}$. The standard deviation of the normal distribution is $\lambda = 0.1$. We perform the experiment 100 times for each parameter setting.

Table 2.3 reports the radial MSE values from a testing set of n observations lying on the true function.

Table 2.3: Experiment 1: Values of the radial MSE relative to the true function. The angle used in CNLS-d estimator varies and the noise direction is randomly selected. In the DGP, the standard deviation of the noise distribution, λ, is 0.1.

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

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

As shown in Table 2.3, the angle corresponding to the smallest MSE (shown in bold) is the one that gives an orthogonal direction to the center of the true function, $\frac{\pi}{4}$, and that the MSE values differ significantly, increasing at similar rates as the direction angle deviates from $\frac{\pi}{4}$ in either direction.

2.5.2.2 Experiment 2 - The base case with fixed noise directions

In this experiment, θ_{ϵ_i}, which characterizes the noise direction for each observation, is constant for all observations, θ_ϵ. The values used for θ_ϵ and the directions in CNLS-d estimator are the same, $0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}$. The standard deviation of the normal distribution is again $\lambda = 0.1$. We perform the experiment 100 times for each parameter settings. Table 2.4 reports the results.

Each row in the Table 2.4 corresponds to a different noise direction in DGP. The bold numbers identify the directions in the CNLS-d estimator that obtain the smallest MSE for each noise direction. We confirm our previous insight, from the parametric estimator and fixed noise direction case described in Appendix A.2.1, that the bold values appearing on the diagonal (from the upper-left to the lower-right of Table 2.4) correspond to the directions used in CNLS-d. This result indicates

that selecting the direction in the SDDF that matches the underlying noise direction in the DGP results in improved functional estimates.

Table 2.4: Experiment 2: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d estimator direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.1.

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

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

2.5.2.3 Experiment 3. Base case with fixed noise direction and different noise levels

In Experiment 3, we vary the noise term by changing the λ coefficient. Table 2.5 reports the results for $\lambda = 0.05$.

Table 2.5: Experiment 3–Less Noise: Values of radial MSE relative to the true function varying the DGP noise direction and the CNLS-d direction. In the DGP, the standard deviation of the noise distribution, λ, is 0.05.

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

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

In Table 2.5 (Experiment 3, with $\lambda = 0.05$), we do not observe the same diagonal pattern observed in Experiment 2, and the best direction for CNLS-d estimator does not match the direction selected for the noise. This leads us to hypothesize that when the noise level is small, data characteristics, such as the distribution of the regressors or the shape of the function, affect the estimation whereas when the noise level is large, regressors' relative variability becomes a more dominant factor in determining the best direction for the CNSL-d estimator.

However, with $\lambda = 0.2$ the results of Experiment 3 are consistent with those from Experiment 2; i.e., the best direction always coincides with the noise direction selected. The results of Experiment 3 with $\lambda = 0.2$ are reported in Appendix A.2, Table A.3 (Experiment 3 with $\lambda = 0.2$).

2.5.2.4 Experiment 4: Base case with different distributions for the initial observations on the true function

In Experiment 4, we seek to understand how changing the DGP for the angle, $\theta_i,\ i = 1, \ldots, n$, affects the optimal direction. We consider the three normal distributions with different parameters: $N\left[\frac{\pi}{8}, \frac{\pi}{16}\right]$, $N\left[\frac{\pi}{4}, \frac{\pi}{16}\right]$ and $N\left[\frac{3\pi}{8}, \frac{\pi}{16}\right]$. We truncate the tails of the distribution so that the generated angles fall in the range $[0, \pi/2]$. Noise is specified as in Experiment 1. Table 2.6 reports the results of this experiment.

Table 2.6: Experiment 4: Values of radial MSE relative to the true function varying the CNLS-d direction and the mean of the normal distribution used in the DGP.

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

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

In Table 2.6, we observe that selecting a direction in the SDDF to match $\bar{\theta}$, the mean of the distribution for the angle variable used in the DGP, corresponds to the smallest MSE value. This

result suggests that the estimator's performance improves when we select a direction that points to the "center" of the data.

Appendix A.2.2 presents additional experiments, varying the distribution of the observations and considering three outputs with a fixed costed level. These experiments lend further support to the strategy of selecting a direction pointed to the "center" of the data.

2.6 Proposed Approach to Direction Selection

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

1. Normalize the data:

$$\breve{y}_{ij} = \frac{y_{ij} - \min_k y_{kj}}{\max_k y_{kj} - \min_k y_{kj}},\ j = 1, \ldots, Q,\ i, k = 1, \ldots, n \tag{2.24}$$

$$\breve{c}_i = \frac{y_i - \min_k c_k}{\max_k c_k - \min_k c_k},\ i, k = 1, \ldots, n \tag{2.25}$$

2. Select the direction:

$$\begin{bmatrix} g^{y_1} \\ \vdots \\ g^{y_Q} \\ g^{c} \end{bmatrix} = \begin{bmatrix} \text{median}\,(\breve{y}_1) \\ \vdots \\ \text{median}\,(\breve{y}_Q) \\ 1 - \text{median}\,(\breve{c})\,. \end{bmatrix} \tag{2.26}$$

This provides a method for direction selection that can be used in applications when the true direction is unknown.[13] We test the proposed method by estimating a cost function for a US

[13] A cost function is convex with respect to the point $[\breve{y}_1, , \breve{y}_Q, \breve{C}] = [0, , 0, 1]$. Therefore, to have a ray that

hospital data set.

2.7 Cost Function Estimation of the US Hospital Sector

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

2.7.1 Description of the Data Set

We obtain cost data from the American Hospital Association's (AHA) Annual Survey Databases from 2007 to 2009. The costs reported include payroll, employee benefits, depreciation, interest, supply expenses and other expenses. We estimate a cost function which can be interpreted as a distance function with a single input when hospitals face the same input prices[14]. We obtain hospital output data from the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) core file that captures data annually for all discharges for a 20% sample of US community hospitals. The hospital sample changes every year. For each patient discharged, all procedures received are recorded as International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9-CM) codes. The typical hospital in the US relies on these detailed codes to quantify the medical services it provides (Zuckerman et al. (1994)). We map the codes to four categories of procedures, specifically the procedure categories are "Minor Diagnostic," "Minor Therapeutic," "Major Diagnostic," and "Major Therapeutic" which are standard output categories in the literature (Pope and Johnson (2013)). The number of procedures is each category are summed for each hospital by year to construct the output variables. The total number of hospitals sampled is around 1,000 per year from 2007 to 2009.[15] However, mapping between the two databases is only possible

points from the point $[0,,0,1]$ to the median of the data, the directional vector $[\text{median}(\breve{y}_{i1}),...,\text{median}(\breve{y}_{iQ}),1-\text{median}(\breve{c}_i)]$ is needed. Alternatively for a production function we would use the center point $[\breve{x}_1,,\breve{x}_d,\breve{y}] = [1,,1,0]$ and construct a ray from this point to the median of the data resulting in the directional vector $[1-\text{median}(\breve{x}_{i1}),...,\text{median}(\breve{x}_{id}),\text{median}(\breve{y}_i)]$.

[14]Unfortunately we do not observe input prices. We chose to estimate a cost function and make the assumption of common input prices rather than impose an arbitrary division of the cost.

[15]The NIS survey is a stratified systematic random sample. The strata criteria are urban or rural location, teaching status, ownership, and bed size. This stratification ensures a more representative sample of discharges than a sim-

for approximately 50% of the hospitals in the HCUP data, resulting in approximately 450 to 525 observations available each year.

Table 2.7: Summary Statistics of the Hospital Data Set

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

2.7.2 Pre-Analysis of the Data Set

2.7.2.1 Testing the Relevance of the Regressors

We begin by testing the statistical significance of our four output variables, $\boldsymbol{y} = (y_1, y_2, y_3, y_4)$, for predicting cost. While the variables selected have been used in previous studies, we use these

ple random sample would yield. For details see on `https://www.hcup-us.ahrq.gov/HCUP_Overview/HCUP_Overview/index508_2018.jsp`

tests to evaluate whether this variable specification can be rejected for the current data set of U.S. hospitals from 2007-2009.

The null hypothesis stated for the qth output is:

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

against:[16]

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

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

2.7.3 Results

2.7.3.1 CNLS-d and Different Directions

We analyze each year of data as a separate cross-section because, as noted above, the HCUP does not track the same set of hospitals across years. To illuminate the direction's effect on the functional estimates, we graph "Cost" as a function of "Major Diagnostic Procedures" and "Major Therapeutic Procedures" holding "Minor Diagnostic Procedures" and "Minor Therapeutic Procedures" constant at their median values. Figure 2.6 illustrates the estimates for three different directions, one with only a cost component, one with only a component in Major Therapeutic Procedures, and one that comes from our median approach. Visual inspection indicates that the estimates with different directions produce significantly different estimates, highlighting the importance of considering the question of direction selection.

We compare the estimator's performance when using different directions. Table 2.8 reports the MSE for three sample directions in each year. We define our direction vector as $(g^{y1}, g^{y2}, g^{y3}, g^{y4}, g^{c})$.[17]

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

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

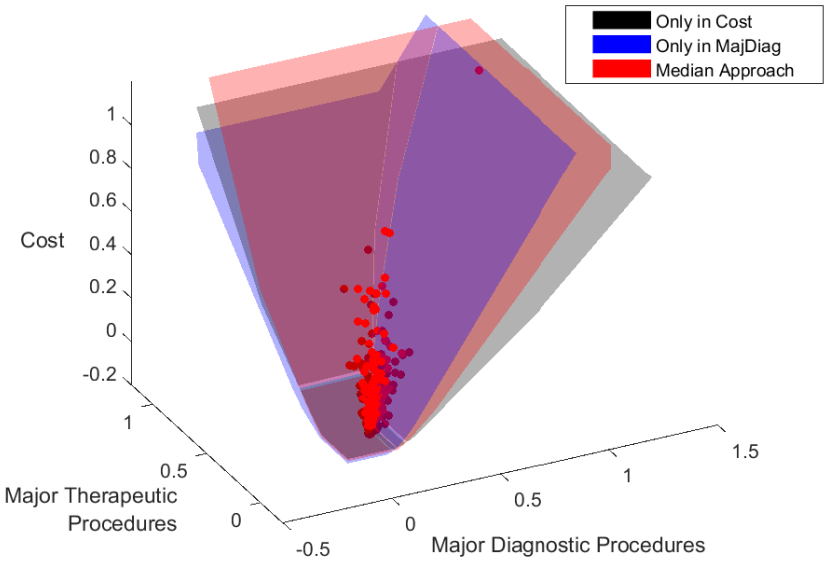

Figure 2.6: US Hospital Cost Function Estimates for Three Directions

Table 2.8: Results of the radial MSE values for different directions by year

Direction $(g^{y_1}, g^{y_2}, g^{y_3}, g^{y_4}, g^c)$	Year 2007	2008	2009
(0.45, 0.45, 0.45, 0.45, 0.45)	2.10	**1.30**	1.50
(0.35, 0.35, 0.35, 0.35, 0.71)	2.15	1.65	**1.29**
Median Direction	**1.79**	1.55	1.34

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

We pick two directions, one with equal components in all dimensions, and a second direction that has a cost component that is double the value of the output components. The median vector is $(0.014, 0.041, 0.033, 0.038, 0.998)$, which is very close to the cost-only direction. The MSE varies by 15-30% over the different directions. We observe that there is no clear dominant direction; however, the median direction performs reasonably well in all cases. We conclude that as long as a direction with non-zero components for all variables that could contain noise is selected, then the

precise direction selected is not critical to obtaining improved estimation results.

2.7.3.2 Comparison with other estimators

We compare three methods to estimate a cost function: 1) a quadratic functional form (without the cross-product terms), Färe et al. (2010); 2) CNLS-d with the direction selection method proposed in Section 2.6; and 3) lower bound estimate calculated using a local linear kernel regression with a Gaussian kernel and leave one-out cross-validation for bandwidth selection, Li and Racine (2007).[18] We select these estimators because a quadratic functional form to model production has been used in recent productivity and efficiency analysis of healthcare. See, for example, Ferrier et al. (2018). The local linear kernel is selected because it is an extremely flexible nonparametric estimator and provides a lower bound for the performance of a functional estimate. However, note that the local linear kernel does not satisfy standard properties of a cost function; i.e., cost is monotonic in output and marginal costs are increasing as output increases.

We will use the criteria of K-fold average MSE with $k = 5$ to compare the approaches. This means we split the data equally into 5 parts. We use 4 of the 5 parts for estimation (training) and evaluate the performance of the estimator on the 5th part (testing). We do this for all 5 parts and average the results. The values presented in Table 2.9 correspond to the average across folds.

Table 2.9: US Hospital K-fold Average MSE in Cost to the Cost Function Estimates for the Three Functional Specifications by Year

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

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

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

While the average MSEs for all years are lowest for the lower bound estimator, CNLS-d performs relatively well as it is close to the lower bound in terms of fitting performance while imposing standard axioms of a cost function. As is true of most production data, the hospital data are very noisy. The shape restrictions imposed in CNLS-d improves the interpretability. The CNLS-d estimator outperforms the parametric approach, indicating the general benefits of nonparametric estimators.

2.7.3.3 Description of Functional Estimates - MPSS and Marginal Costs

We report the most productive scale size (MPSS) and the marginal costs for the a quadratic parametric estimator, the CNLS-d estimator with our proposed direction selection method, and an alternative.[19] These metrics are determined on the averaged K-fold estimations for each estimation method. For the MPSS, we present the cost levels obtained for different ratios of *Minor Therapeutic* procedures (MinTher) and *Major Therapeutic* procedures (MajTher), with the minor and major diagnostics held constant at their median levels.

MPSS results are presented in Table 2.10 and the values for CNLS-d (Median Direction) are illustrated in Figure 2.7. We observe small variations across both years and estimators. The differences across years are in part due to the sample changing across years. Most hospitals are small and operate close to the MPSS. However, there are several large hospitals that are operating significantly above MPSS. Hospitals might choose to operate at larger scales and provide a large array of services allowing consumers to fulfill multiple healthcare needs.

For marginal costs, we present the values for different percentiles of the MinTher and MajTher, with the minor and major diagnostics held constant at their median levels. A more exhaustive comparison across all outputs is presented in Appendix A.3. Marginal cost information can be used by hospital decision makers to select the types of improvements that are likely to result in higher productivity with minimal cost increase. For example, consider a hospital that is in the 50^{th} percentile of the data set for all four outputs in 2008 and the hospital manager has the option

[19]Here most productive scale size is measured on each ray from the origin (fixing the output ratios) and is defined as the cost level that maximizes the ratio of aggregate output to cost. Marginal cost is measured on each ray from the origin (fixing the output ratios) and is defined as the cost to increase aggregate output by one unit.

to expand operations for either minor or major diagnostic procedures. Results reported in Tables 2.11 and 2.12 indicate that an increase of 1 minor therapeutic procedures would result in a $\$4.9k$ increase in cost. Alternatively, an increase of 1 major therapeutic procedures would result in a $\$7.7k$ increase in cost. A decision maker would want to consider the revenue generated by the different procedures; however, these estimates provide insights regarding the incremental cost of additional major and minor therapeutic procedures.

CNLS-d is the most flexible of the estimators and allows MPSS values to fluctuate significantly across percentiles. CNLS-d does not smooth variation, rather it minimizes the distance from each observation to the shape constrained estimator. In Appendix A.3, results for the local linear kernel estimator are also presented. Even though the local linear kernel bandwidths are selected via cross-validation, relatively large values are selected due to the relatively noisy data and the highly skewed distribution of output. These large bandwidths and the parametric nature of the quadratic function make these two estimators relatively less flexible compared to CNLS-d. A feature of performance that is captured only by CNLS-d is that, hospitals specializing in either minor or major therapeutics maximize productivity at a larger scales of operation as illustrated in Figure 2.7.

Table 2.10: Most Productive Scale Size measured in cost ($\$M$) conditional on Minor Therapeutic procedures (MinTher) and Major Therapeutic procedures (MajTher), Minor Diagnostic procedures (MinDiag) and Major Diagnostic procedures (MajDiag) held constant at the 50th percentile

Ratio	Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MajTher/MinTher	2007	2008	2009	2007	2008	2009	2007	2008	2009
20%	13	379	252	210	61	88	224	137	106
30%	17	861	640	146	66	83	134	129	148
40%	272	377	1090	107	56	77	127	85	135
50%	870	249	1552	112	64	85	124	126	134
60%	360	210	276	90	70	120	88	96	142
70%	205	182	187	111	66	184	132	104	104
80%	151	170	150	174	69	286	221	110	111

Note: The values displayed are in $M

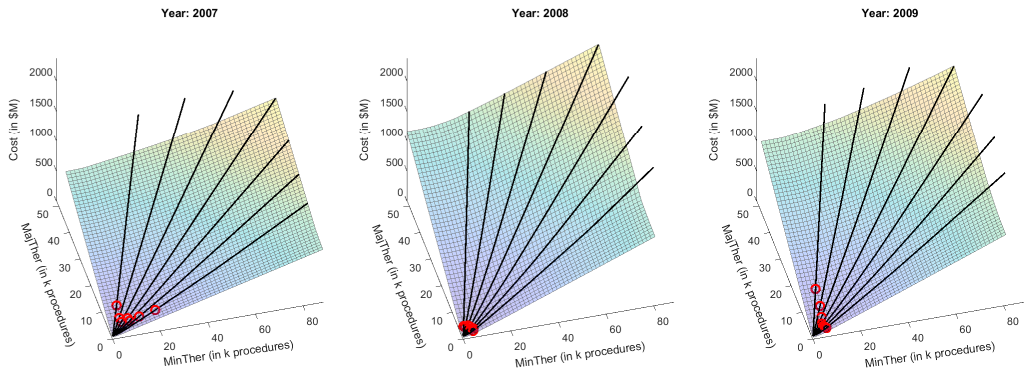

Figure 2.7: Most Productive Scale Size (in red) on the estimated function by "CNLS-d (Med)", CNLS-d using the median approach for the direction, for different ratios of Major Therapeutic Procedures over Minor Therapeutic Procedures

Table 2.11: Marginal Cost of Minor Therapeutic Procedures

Percentile		Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MinTher	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	8.9	6.5	13.2	0.03	0.03	0.03	0.2	0.02	0.1
25	50	8.9	6.5	13.2	0.05	0.1	0.1	0.04	0.1	0.04
25	75	8.9	6.5	13.2	0.2	0.04	0.03	0.1	0.02	0.02
50	25	8.1	6.1	12.4	6.9	5.5	7.4	5.9	6.3	7.8
50	50	8.1	6.1	12.4	4.3	4.9	7.8	2.1	3.7	7.4
50	75	8.1	6.1	12.4	0.2	0.4	0.03	0.1	0.02	0.02
75	25	6.0	5.0	10.4	9.6	13.5	14.0	9.5	10.9	14.1
75	50	6.0	5.0	10.4	9.6	13.5	14.3	9.6	10.9	13.8
75	75	6.0	5.0	10.4	5.7	10.1	6.4	4.6	8.7	6.4

Note: The values displayed are in $k

The marginal cost results for Minor Therapeutic procedures are presented in Table 2.11 and Figure 2.8 (left) and the marginal cost results for Major Therapeutic procedures are reported in Table 2.12 and Figure 2.8 (right). As was the case for MPSS (see Table 2.10), CNLS-d is more flexible and its marginal cost estimates vary significantly across percentiles. The CNLS-d with different directions provides very similar marginal costs estimates. However, the CNLS-d estimates differ significantly from the marginal cost estimates obtained with the parametric estimator. For

Table 2.12: Marginal Cost of Major Therapeutic Procedures

Percentile		Quadratic Regression			CNLS-d (median)			CNLS-d (equal)		
MinTher	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	10.5	11.5	9.8	0.1	0.04	0.1	0.2	0.03	0.1
25	50	11.7	13.0	10.8	11.3	11.8	15.7	10.5	10.3	14.6
25	75	15.1	17.2	14.5	19.8	22.1	24.6	19.8	21.8	24.0
50	25	10.5	11.5	9.8	0.4	0.2	0.5	0.1	0.1	0.4
50	50	11.7	13.0	10.8	3.7	7.7	1.7	6.9	7.1	3.7
50	75	15.1	17.2	14.5	19.8	22.0	24.6	19.8	21.8	24.0
75	25	10.5	11.5	9.8	0.2	0.03	0.1	0.0	0.1	0.1
75	50	11.7	13.0	10.8	0.2	0.2	0.4	0.8	0.1	0.3
75	75	15.1	17.2	14.5	18.3	12.4	19.8	16.2	11.0	15.2

Note: The values displayed are in $k

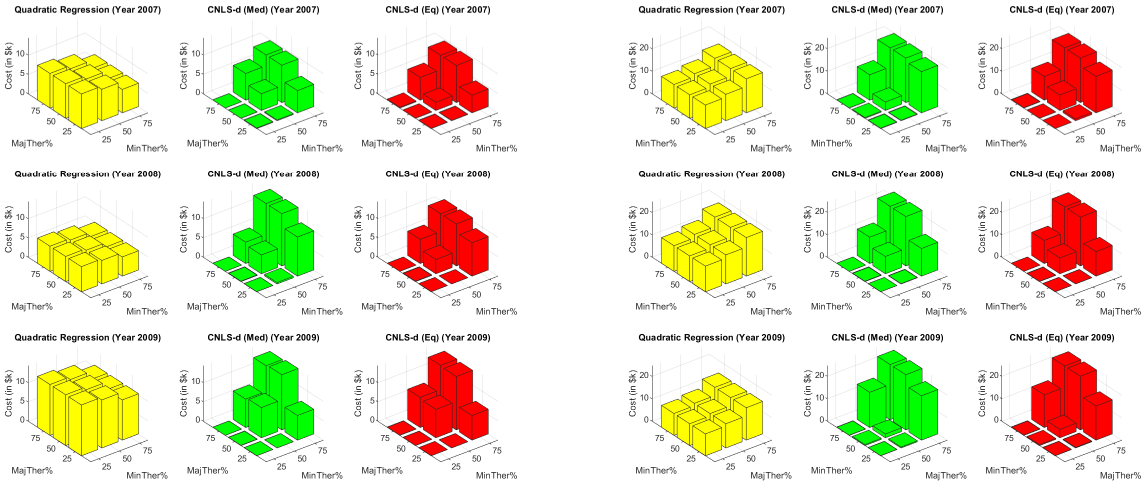

Figure 2.8: Marginal Cost of the Minor Therapeutic procedures (left) and Marginal Cost of the Major Therapeutic procedures (right) ("CNLS-d (Med)" corresponds to CNLS-d using the median approach for the direction and "CNLS-d (Eq)" corresponds to CNLS-d using the direction with equal components in all netputs

CNLS-d the marginal costs results are in line with the theory that marginal costs are increasing with scale. This property can also be violated if using a non-parametric estimator without any shape constraints imposed. For example this can be seen in the marginal costs of minor therapeutic

procedures for the parametric (quadratic) regression estimator, Figure 2.8.

Our data set, which combines AHA cost data with AHRQ output data for a broad sample of hospitals from across the US, is unique to the best of our knowledge. However, the marginal cost estimates are broadly in line with marginal cost estimates for US hospitals for similar time periods. Gowrisankaran et al. (2015) studied a considerably smaller set of Northern Virginia hospitals observed in 2006 that, on average, were larger that hospitals in our data set. Due to the differences in the measures of output the marginal cost levels are not directly comparable. However, conditional on the size variation, the variation in marginal costs is similar to the variation we observe for the parametric (quadratic) regression specification applied to our data. Boussemart et al. (2015) analyzed data on nearly 150 hospitals located in Florida observed in 2005. The authors use a different output specification and a translog model; however, their distribution of hospital size is similar to our data set and we observe similar variances in marginal costs with the parametric (quadratic) regression specification applied to our data.

2.8 Conclusions

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

In future research, we propose developing an alternative estimator that incorporates multiple directions in CNLS-d while maintaining the concavity axiom. This would permit treating subgroups within the data, allowing different assumptions to be made across subgroups (e.g., for-profit vs. not-for-profit hospitals).

3. APPROXIMATION TO SHAPE CONSTRAINTS OF NON-PARAMETRIC ESTIMATORS

3.1 Introduction

Production economics theory states some assumptions concerning the shape of a cost function. When using a non-parametric estimator with the adequate shape constraints, the user benefits from a gain in interpretability compare to a fully non-parametric model and improved finite sample performance.

In this chapter, we study several approximations to economically motivated shape constraints that are common in the non-parametric estimator literature and we evaluate their respective performance. The performance is evaluated based on three criteria: goodness of fit, non-violation of the shape constraints and computational feasibility. In this chapter, we focus on the cost function and its related shape constraints, the function being non-decreasing with a decreasing return to scale, leading to convexity. Concerning the approximations, two of them are studied: using a piece-wise linear estimator instead of a smoother estimator and using coordinate-wise constraints instead of global constraints.

The first approximation regarding smoothness is used in estimators like CNLS presented in Kuosmanen (2008), MBCR in Hannah and Dunson (2013) and SCKLS in Yagi et al. (2018). The estimated function is an envelope of hyperplanes, a piece-wise linear representation that is particularly convenient to impose the desired shape constraints. However since the representation is piece-wise linear, the marginal costs, that are associated with the values of the partial derivatives, are constant for some regions of the function's domain and are discontinuous. This is a problem for some users who would prefer a smooth function with continuous marginal costs.

The second approximation is about the way the shape constraints are imposed, specifically coordinate-wise constraints are used rather than global constraints. This means that for instance the control is on the sign of each first and second partial derivative as in the spline estimators presented in Pya and Wood (2015). However this is not sufficient, economic theory implies that

the gradient (i.e. in the cost function case the marginal cost), should be non-negative and increasing for any increase in a regressor or combination of regressors. This is what we call global convexity, while coordinate-wise convexity only enforces that conditional on holding all but one regressor fixed, expanding a single regressor increases the value of the gradient.

We seek to understand the impact of each of these assumptions separately. For the smooth estimators, we develop and evaluate B-spline estimators with shape constraints. These B-splines estimators are based on the work of Pya and Wood (2015), that introduced penalized B-splines with shape constraints. For the piece-wise linear estimators, we extend the SCKLS estimator introduced in Yagi et al. (2018).

Additionally we study the way the shape constraints are imposed. The necessary and sufficient conditions for convexity are enforced on a set of control points (in CNLS this is the observations or in SCKLS this is a set of grid points). CNLS and SCKLS assure global convexity by using linear interpolation between the control points, specifically the final estimated function corresponds to the upper envelope of the estimated hyperplanes in the case of the cost function (lower envelope in the case of the production function). Du et al. (2013) and Yagi et al. (2018)[1] present a Kernel regression model that imposes convexity at as set of control points for smooth estimators. As the number of control points becomes dense on the domain of the function, global convexity is imposed. Here we extend this concept to a spline estimator with shape constraints on a set of control points which we refer to as “Spline Estimator w/ local Afriat constraints". Table 3.1 summarizes the four estimators considered. The estimators in *italics* are developed in this chapter.

The chapter is organized as follows. Section 3.2 gives more information on production economics theory and the shape constraints that are desired. Section 3.3 describes the estimators used for the remainder of the chapter, in Section 3.4 where we test these estimators on Monte-Carlo experiments, and in Section 3.5 where we apply them to US hospitals data. Section 3.6 concludes.

[1] See the appendix of Yagi et al. (2018).

Table 3.1: The four estimators considered to establish a comparison of the different approximations to the shape constraints. The estimators in italics are created for this analysis.

		Smoothness	
		Not smooth	Smooth
Shape restrictions	Coordinate-wise	*SCKLS w/ coordinate-wise constraints*	Spline Estimator derived from SCAM
	Global	SCKLS	*Spline Estimator w/ local Afriat constraints*

3.2 Production Economics Theory and Shape Constraints

3.2.1 Axioms of the Cost Function

Let f the cost function such that $c = f(\boldsymbol{y})$, where c is a scalar representing the cost and $\boldsymbol{y}$ is a $\mathbf{R}^d$ vector representing the outputs.

$f : \mathbf{R}_+^d \to \mathbf{R}$ and the two main properties assumed are:

1. **Non-decreasing in $\boldsymbol{y}$**: if $\forall \boldsymbol{y}^0 \in \mathbf{R}_+^d, \forall \boldsymbol{y}^1 \in \mathbf{R}_+^d, \boldsymbol{y}^1 \geq \boldsymbol{y}^0$ then $f(\boldsymbol{y}^1) \geq f(\boldsymbol{y}^0)$

2. **Convexity in $\boldsymbol{y}$**: $\forall \theta \in [0,1], \forall \boldsymbol{y}^0 \in \mathbf{R}_+^d, \forall \boldsymbol{y}^1 \in \mathbf{R}_+^d, f\left(\theta \boldsymbol{y}^0 + (1-\theta)\boldsymbol{y}^1\right) \leq \theta f(\boldsymbol{y}^0) + (1-\theta) f(\boldsymbol{y}^1)$

These two main properties are justified by the economics theory:

1. **Non-decreasing in $\boldsymbol{y}$**: It represents that the more outputs are generated the cost should increase.

2. **Convexity in $\boldsymbol{y}$**: The function is convex which corresponds to an increasing return to scale, the more you produce the more it becomes difficult to use your inputs efficiently to produce more outputs. This represents increasing marginal cost along the corresponding output.

3.2.2 Coordinate-wise constraints versus global constraints

The coordinate-wise constraints are defined as constraints on the signs of the first and second partial derivatives. For a function $f : \mathbf{R}_+^d \to \mathbf{R}$:

1. **Non-decreasing in $\boldsymbol{y}$**: $\frac{\partial f}{\partial y_k}(\boldsymbol{y}) \geq 0,$ for all $k = 1, \ldots, d,$ for all $y \in \mathbf{R}_+^d$

2. **Convexity in $\boldsymbol{y}$**: $\frac{\partial^2 f}{\partial y_k^2}(\boldsymbol{y}) \geq 0,$ for all $k = 1, \ldots, d,$ for all $y \in \mathbf{R}_+^d,$

It is easy to find counterexamples to show that these conditions are not equivalent to the global constraints. Still it can be challenging to impose such constraints depending on the functional representation used by the estimator. The estimators using coordinate-wise constraints introduced in section 3.3 will be either imposing sufficient conditions or local enforcement of these constraints.

3.2.3 Afriat Inequalities and constraints validation

In order to verify the validity of the global shape constraints specified above, the Afriat inequalities are tested on a set of "control points". These inequalities are still written in the case of the cost function, thus for shape constraints that are convexity and non-decreasing in $\boldsymbol{y}$:

$$\hat{c}_i - \hat{c}_j \geq \boldsymbol{b_j'}\left(\boldsymbol{y}_i - \boldsymbol{y}_j\right), \qquad \text{for all } i, j = 1, \ldots, n,\ i \neq j \tag{3.1a}$$

$$\boldsymbol{b_i} \geq 0, \qquad \text{for all } i = 1, \ldots, n \tag{3.1b}$$

where $\hat{c}_i$ is the evaluation of the estimated function at the control output level $\boldsymbol{y}_i$, and $\boldsymbol{b_i}$ is the numerical approximation of the gradient of the estimated function at $\boldsymbol{y}_i$. Condition (3.1a) imposes convexity while condition (3.1b) enforces monotonicity. These are not only used to check the shape of the function but also in the notations of some estimators.

3.3 Estimators

3.3.1 Shape Constrained Kernel-weighted Least Squares (SCKLS)

3.3.1.1 Regular SCKLS

Shape Constrained Non-parametric Least-Squares (SCKLS) is introduced in Yagi et al. (2018). The formulation presented here is the cost function estimation case, thus with non-decreasing monotonicity and global convexity constraints enforced. To be more specific, SCKLS with a local linear kernel approximation is used. The formulation is the following:

$$
\begin{aligned}
\min_{a_i, \boldsymbol{b_i}} \quad & \sum_{i=1}^{m}\sum_{j=1}^{n}(c_j - a_i - (\boldsymbol{Y_j} - \boldsymbol{y_i})'\boldsymbol{b_i})^2 K\left(\frac{\boldsymbol{Y_j} - \boldsymbol{y_i}}{\boldsymbol{h}}\right) & \\
\text{subject to} \quad & a_i - a_l \geq \boldsymbol{b_i'}(\boldsymbol{y_i} - \boldsymbol{y_l}), & i, l = 1, \ldots, m \\
& \boldsymbol{b_i} \geq 0, & i = 1, \ldots, m
\end{aligned}
\tag{3.2}
$$

where $\boldsymbol{Y_j},\ c_j,\ j = 1, \ldots, n$ are the observation output levels and cost level respectively, while $\boldsymbol{y_i},\ i = 1, \ldots, m$ are a set of points on which the hyperplanes are defined, $a_i,\ i = 1, \ldots, m$ being the respective functional estimates and $\boldsymbol{b_i},\ i = 1, \ldots, m$ the estimated slopes. In Yagi et al. (2018), the most standard way to define the set of points on which the hyperplanes are defined is to use a uniform grid.

Also, $K\left(\frac{\boldsymbol{Y_j} - \boldsymbol{y_i}}{\boldsymbol{h}}\right)$ is a Kernel product and $\boldsymbol{h}$ the vector of bandwidths. The type of kernel function used for the Monte Carlo simulations in section 3.4 is the Gaussian kernel and the approach selected to determine the vector of bandwidths is leave-one-out cross validation (LOOCV) applied to the unconstrained Local Linear Kernel regression estimator. For more information on Kernel regressions see Racine and Li (2004), and Yagi et al. (2018) for bandwidth selection.

The constraints in the formulation of the regular SCKLS estimator are the Afriat inequalities defined in section 3.2.3 and adapted to the piece-wise linear representation that is the SCKLS estimated function. This enforces everywhere the global shape constraints, as the estimated function is an upper envelope of hyperplanes in the convex case considered.

3.3.1.2 SCKLS with coordinate-wise constraints

This is the version of SCKLS for which the shape constraints are modified to correspond to a coordinate-wise approximation. Let $\mathbf{y} = \{\boldsymbol{y_i} \in \mathbf{R}_+^d,\ i = 1, \ldots, m\}$ the set of control point levels. Let $\mathbf{s} = \{s(i,k),\ i = 1, \ldots, m,\ k = 1, \ldots, d \mid y_{s(i,k)} < y_{s(i+1,k)},\ i = 1, \ldots, (m-1),\ k = 1, \ldots, d\}$ the set of ordered indices for each dimension. In the case of a non-decreasing convex cost function, the formulation is:

$$\min_{a_i, \boldsymbol{b_i}} \quad \sum_{i=1}^{m}\sum_{j=1}^{n}(c_j - a_i - (\boldsymbol{Y_j} - \boldsymbol{y_i})'\boldsymbol{b_i})^2 K\left(\frac{\boldsymbol{Y_j} - \boldsymbol{y_i}}{\boldsymbol{h}}\right) \tag{3.3}$$

$$\text{subject to} \quad \frac{a_{s(i+1,k)} - a_{s(i,k)}}{y_{s(i+1,k)} - y_{s(i,k)}} \leq \frac{a_{s(i+2,k)} - a_{s(i+1,k)}}{y_{s(i+2,k)} - y_{s(i+1,k)}}, \qquad i = 1, \ldots, (m-2)\ ,\ k = 1, \ldots, d \tag{3.3a}$$

$$a_{s(2,k)} - a_{s(1,k)} \geq 0 \qquad k = 1, \ldots, d \tag{3.3b}$$

$$a_{s(m,k)} - a_{s(m-1,k)} \geq 0 \qquad k = 1, \ldots, d \tag{3.3c}$$

$$\boldsymbol{b_i} \geq 0, \qquad i = 1, \ldots, m \tag{3.3d}$$

where as in Equation (3.2), $\boldsymbol{Y_j},\ c_j,\ j = 1, \ldots, n$ are the observation output levels and cost level respectively, $\boldsymbol{y_i},\ i = 1, \ldots, m$ are a set of points on which the hyperplanes are defined, $a_i,\ i = 1, \ldots, m$ being the respective functional estimates and $\boldsymbol{b_i},\ i = 1, \ldots, m$ the estimated slopes. $K\left(\frac{\boldsymbol{Y_j} - \boldsymbol{y_i}}{\boldsymbol{h}}\right)$ is the Kernel product and $\boldsymbol{h}$ is the vector of bandwidths as described in section 3.3.1.1. Inequalities (3.3a) are the constraints imposing coordinate-wise convexity, inequalities (3.3b) and (3.3c) impose coordinate-wise monotonicity, and inequalities (3.3d) impose a positive gradient for the hyperplane.

The estimated function can be evaluated at any level using the unconstrained Local Linear Kernel regression as a way to interpolate in between the estimated levels at the control points. The type of kernel function and the vector of bandwidths used for this interpolation are the same as the ones used for the estimation of the coefficients in equation (3.3).

3.3.2 Shape Constrained B-Splines Least-Squares (SCBLS)

3.3.2.1 B-splines introduction and univariate case

B-splines are thoroughly introduced in De Boor et al. (1978). The B-spline of order $(r+1)^{th}$ is expressed as:

$$m(x) = \sum_{j=1}^{q} \gamma_j B_j^r(x) \tag{3.4}$$

where q is the number of basis functions, γ_j are the spline coefficients and B_j^r are the B-spline basis functions determined by the Cox-de-Boor recursion formula. The recursion formula depends on a vector of knots

$\boldsymbol{t} \in \mathbf{R}^{q+r+2}$ and should be evaluated for $x \in [t_{r+2}, t_{q+1}]$:

$$B_j^r(x) = \frac{x - t_j}{t_{j+r+1} - t_j} B_j^{r-1}(x) + \frac{t_{j+r+2} - x}{t_{j+r+2} - t_{j+1}} B_{j+1}^{r-1}(x), \; j = 1, \ldots, q$$
$$B_j^{-1}(x) = \begin{cases} 1, & t_j \leq x < t_{j+1} \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

Here we choose to focus on a sequence of equally spaced knots but the only assumption required about the knots is that $t_{j+1} \geq t_j, \;$ for all $j = 1, \ldots, (q + r + 2), .$

3.3.2.2 Expansion of B-splines to multivariate case

The expansion of B-splines to the multivariate case is through the tensor formulation. For a case with d variables:

$$m(\boldsymbol{x}) = \sum_{j_1=1}^{q_1} \ldots \sum_{j_d=1}^{q_d} \gamma_{j_1 \ldots j_d} B_{j_1 \ldots j_d}^r(\boldsymbol{x}) \tag{3.6}$$

where $B_{j_1 \ldots j_d}^r(\boldsymbol{x}) = \prod_{k=1}^d B_{j_k}^r(x_k)$ is the product of the basis functions for each dimension and $\gamma_{j_1 \ldots j_d} \in \mathbf{R}$ is the spline coefficient. This means that we define d vectors of knots $\boldsymbol{t}_k \in \mathbf{R}^{q_k+r+2}$, one for each dimension.

3.3.2.3 SCBLS with coordinate-wise shape constraints

In this section we detail the first Shape-Constrained B-spline Least Squares (SCBLS) estimator. Our estimator imposes coordinate-wise shape constraints and the estimator is a restricted case of Pya and Wood (2015)'s estimator.

We first write the formulation in the univariate case, as it is easier to read the constraints. We consider a vector of knots $\boldsymbol{t}$ that is uniformly spaced[2] with a given number of basis functions, q, and $(r + 1)$ the order of the B-spline considered. In the univariate case, the formulation for SCBLS with coordinate-wise shape constraints is:

[2]The formulation for a non-uniform grid is introduced in Appendix B.1.

$$\min_{\gamma_j} \quad \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{q} \gamma_j B_j^r(Y_i) \right)^2 \tag{3.7}$$

$$\text{subject to} \quad \gamma_j \geq \gamma_{j-1}, \qquad j = 2, \ldots, q \tag{3.7a}$$

$$\gamma_j - 2\gamma_{j-1} + \gamma_{j-2} \geq 0, \qquad j = 3, \ldots, q \tag{3.7b}$$

where $\boldsymbol{Y_i},\ c_i,\ i = 1, \ldots, n$ are the observation output levels and cost level respectively, $\gamma_j,\ j = 1, \ldots, q$ are the spline coefficients and $B_j^r,\ j = 1, \ldots, q$ are the basis functions of order $(r+1)$ based on the knots vector $\boldsymbol{t}$. The constraints only concern the spline coefficients and are sufficient for the coordinate-wise constraints, the constraints on the signs of the first and second partial derivatives presented in section 3.2.2. Thus inequalities (3.7a) are sufficient conditions for the sign of the first derivatives and inequalities (3.7b) for the sign of the second derivatives. Pya and Wood (2015) map these constraints on the spline coefficients to the signs of the partial first and second partial derivatives.

To write the problem in the multivariate case, we use the tensor formulation introduced in section 3.3.2.2:

$$\min_{\gamma_{j_1 \ldots j_d}} \quad \sum_{i=1}^{n} \left(c_i - \sum_{j_1=1}^{q_1} \ldots \sum_{j_d=1}^{q_d} \gamma_{j_1 \ldots j_d} B_{j_1 \ldots j_d}^r(\boldsymbol{Y_i}) \right)^2 \tag{3.8}$$

$$\text{subject to} \quad \gamma_{j_1 \ldots j_k \ldots j_d} \geq \gamma_{j_1 \ldots (j_k - 1) \ldots j_d}, \qquad k = 1, \ldots, d,\ j_k = 2, \ldots, q_k$$
$$l = \{1, \ldots, d\}\text{-}\{k\},\ j_l = 1, \ldots, q_l \tag{3.8a}$$

$$\gamma_{j_1 \ldots j_k \ldots j_d} - 2\gamma_{j_1 \ldots (j_k - 1) \ldots j_d} + \gamma_{j_1 \ldots (j_k - 2) \ldots j_d} \geq 0, \qquad k = 1, \ldots, d,\ j_k = 3, \ldots, q_k$$
$$l = \{1, \ldots, d\}\text{-}\{k\},\ j_l = 1, \ldots, q_l \tag{3.8b}$$

where $\boldsymbol{Y_i},\ c_i,\ i = 1, \ldots, n$ are the observation output levels and cost level respectively, $\gamma_{j_1 \ldots j_d}$ are the spline coefficients and $B_{j_1 \ldots j_d}^r$ are the products of basis functions as described in section 3.3.2.2. Inequalities (3.8a) are sufficient conditions for the sign of the first derivatives and inequalities (3.8b) for the sign of the second derivatives.

However the latest formulation suffers from the curse of dimensionality as the decision variables are the splines coefficients of which the size increases exponentially with the number of dimensions. The optimization problem can be solved using a quadratic solver for low dimensional problems leveraging the fact the constraint matrix is sparse.

3.3.2.4 SCBLS with local Afriat constraints

In a similar way to the previous section we start with the univariate case as it is clearer to read the constraints. Then we write down the multivariate case that uses the tensor formulation. Given a vector of knots $\boldsymbol{t}$, q the number of basis functions and $(r+1)$the order of the B-spline considered, the formulation for SCBLS with local Afriat constraints in the univariate case is:

$$\min_{\gamma_j} \quad \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{q} \gamma_j B_j^r(Y_i) \right)^2 \tag{3.9}$$

$$\text{subject to} \quad \sum_{j=1}^{q} \gamma_j \left[B_j^r(y_i) - B_j^r(y_l) \right] \geq b_l(y_i - y_l), \qquad i, l = 1, \ldots, m,\ i \neq l \tag{3.9a}$$

$$b_i \geq 0, \qquad i = 1, \ldots, m \tag{3.9b}$$

where $Y_i,\ i = 1, \ldots, n$ are the observations output levels, $c_i,\ i = 1, \ldots, n$ are the observations cost levels, while $y_i,\ i = 1, \ldots, m$ are a set of control points where the Afriat inequalities are imposed. b_i is the numerical approximation of the gradient of the estimated function at $\boldsymbol{y}_i$. Specifically if we define the small values used for the numerical approximation as $\epsilon_d = 10^{-10}$,

$$b_i = \frac{\sum_{j=1}^{q} \gamma_j \left[B_j^r(y_i + \epsilon_d) - B_j^r(y_i - \epsilon_d) \right]}{2\epsilon_d}, \quad i = 1, \ldots, m \tag{3.10}$$

Inequalities (3.9a) impose the local Afriat inequalities between each pair of "control points", $\boldsymbol{y}_i,\ i = \ldots, m$, and inequalities (3.9b) impose monotonicity. Thus they enforce the restrictions suggested by economic theory on the set of control points[3].

[3]Note because linear interpolation is not used, the restrictions are only *local* Afriat inequalities.

Using the tensor formulation to extend the estimator for multivariate cases, we can write:

$$\min_{\gamma_{j_1\dots j_d}} \quad \sum_{i=1}^{n}\left(c_i - \sum_{j_1=1}^{q_1}\dots\sum_{j_d=1}^{q_d}\gamma_{j_1\dots j_d}B^r_{j_1\dots j_d}(\boldsymbol{Y_i})\right)^2 \tag{3.11}$$

$$\text{subject to} \quad \sum_{j_1=1}^{q_1}\dots\sum_{j_d=1}^{q_d}\gamma_{j_1\dots j_d}\left[B^r_{j_1\dots j_d}(\boldsymbol{y_i}) - B^r_{j_1\dots j_d}(\boldsymbol{y_l})\right] \geq \boldsymbol{b_l'}(\boldsymbol{y_i}-\boldsymbol{y_l}), \quad i,l=1,\dots,m,\ i\neq l \tag{3.11a}$$

$$\boldsymbol{b_i} \geq \boldsymbol{0}, \qquad i=1,\dots,m \tag{3.11b}$$

where $\boldsymbol{Y_i},\ i=1,\dots,n$ are the observations output levels, $c_i,\ i=1,\dots,n$ are the observations cost levels, while $\boldsymbol{y_i},\ i=1,\dots,m$ are the set of control points for the Afriat inequalities. In the multivariate case, the formulation for $\boldsymbol{b_i}$, the numerical approximation of the gradient of the estimated function at $\boldsymbol{y}_i$ is:

$$b_{ik} = \left\{\sum_{j_1=1}^{q_1}\dots\sum_{j_d=1}^{q_d}\gamma_{j_1\dots j_d}\left[B^r_{j_1\dots j_d}\left(\begin{bmatrix} y_1 \\ \vdots \\ (y_k+\epsilon_d) \\ \vdots \\ y_d \end{bmatrix}\right) - B^r_{j_1\dots j_d}\left(\begin{bmatrix} y_1 \\ \vdots \\ (y_k-\epsilon_d) \\ \vdots \\ y_d \end{bmatrix}\right)\right]\right\} / \left(2\epsilon_d\right), \tag{3.12}$$

$$i=1,\dots,m,\ k=1,\dots,d$$

3.3.2.5 SCBLS with coordinate-wise constaints and local Afriat constraints

This estimator combines the constraints of the two previous introduced spline estimators. It belongs to the sector of Table 3.1 corresponding to "Global" shape restrictions[4]. It is a hybrid version of the two previous estimator that imposes more structure than the two previous estimators, but is similar if not slightly more computationally intensive than the estimator with only local Afriat constraints. Since we already introduced the estimators before we only write the multivariate case.

Given a vector of knots $\boldsymbol{t}$, q the number of basis functions and $(r+1)$ the order of the B-spline consid-

[4]Note the coordinate-wise constraints should be redundant; however, because the constraints used are truly imposing coordinate-wise convexity, meaning coordinate-wise convexity is imposed on any point inside the grid constructed by the set of control points selected, the constraints are more restrictive in some ways than local convexity that only imposed convexity on the control points.

ered, the formulation for SCBLS with coordinate-wise constaints and local Afriat constraints is:

$$\min_{\gamma_{j_1\cdots j_d}} \quad \sum_{i=1}^{n}\left(c_i - \sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\gamma_{j_1\ldots j_d}B^r_{j_1\ldots j_d}(\boldsymbol{Y_i})\right)^2 \tag{3.13}$$

subject to

$$\gamma_{j_1\ldots j_k\ldots j_d} \geq \gamma_{j_1\ldots(j_k-1)\ldots j_d}, \qquad k=1,\ldots,d,\ j_k=2,\ldots,q_k,\quad l=\{1,\ldots,d\}\text{-}\{k\},\ j_l=1,\ldots,q_l \tag{3.13a}$$

$$\gamma_{j_1\ldots j_k\ldots j_d} - 2\gamma_{j_1\ldots(j_k-1)\ldots j_d} + \gamma_{j_1\ldots(j_k-2)\ldots j_d} \geq 0, \qquad k=1,\ldots,d,\ j_k=3,\ldots,q_k,\quad l=\{1,\ldots,d\}\text{-}\{k\},\ j_l=1,\ldots,q_l \tag{3.13b}$$

$$\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=1}^{q_d}\gamma_{j_1\ldots j_d}\left[B^r_{j_1\ldots j_d}(\boldsymbol{y_i}) - B^r_{j_1\ldots j_d}(\boldsymbol{y_l})\right] \geq \boldsymbol{b_l'}(\boldsymbol{y_i}-\boldsymbol{y_l}), \qquad i,l=1,\ldots,m,\ i\neq l \tag{3.13c}$$

$$\boldsymbol{b_i} \geq \boldsymbol{0}, \qquad i=1,\ldots,m, \tag{3.13d}$$

where $\boldsymbol{Y_i},\ i=1,\ldots,n$ are the observations output levels, $c_i,\ i=1,\ldots,n$ are the observations cost levels, while $\boldsymbol{y_i},\ i=1,\ldots,m$ are the set of control points for the Afriat inequalities. $\boldsymbol{b_i}$ is the numerical approximation of the gradient at $\boldsymbol{y_i}$. Inequalities (3.13a) and (3.13b) impose coordinate-wise monotonicity and coordinate-wise convexity, respectively. Inequalities (3.13c) impose the Afriat inequalities and (3.13d) restrict the gradient to be positive at the control points, $\boldsymbol{y_i}$.

3.4 Monte Carlo simulations

In this section, we test the estimators on data generated using a quadratic cost function. The study defines a base case and varies the noise level and the numbers of observations. Additional experiments are performed in which we introduce another approximation, specifically bounding the derivative of the estimated function to improve computational performance. We conclude the section by summarizing the results and giving recommendations based in the insights gained.

3.4.1 Experiments with different data generation settings

3.4.1.1 Experiment 1 - Base case: A Cost Function varying the number of regressors.

A cost function is simulation which defines the relation between cost and outputs in a production process. The cost function we consider satisfies the economic axioms explicitly written and discussed in section 3.2.1, specifically monotonicity and convexity. For the base case we consider several numbers of outputs $d = \{2, 3, 4\}$ and fix the number of observations, $n = 100$. The complete Data Generation Process (DGP) is detailed in Algorithm 1.

Algorithm 1

1. The outputs are drawn from a continuous uniform distribution $y_{ij} \sim U[u_{min}, u_{max}],\ i = 1, \ldots, d,\ j = 1, \ldots, d$, where $u_{min} = 0.01$ and $u_{max} = 0.99$.

2. We define the function f:

$$f \colon \mathbf{R}^d \to \mathbf{R},\ \boldsymbol{y} \mapsto \sum_{k=1}^{d} y_k^2, \tag{3.14}$$

   Let $\boldsymbol{y^{min}} \in \mathbf{R}^d$ such that $y_k^{min} = u_{min},\ k = 1, \ldots, d$ and let $\boldsymbol{y^{max}} \in \mathbf{R}^d$ such that $y_k^{max} = u_{max},\ k = 1, \ldots, d$. The cost values on the true function corresponding to each observation is determined as:

$$\tilde{c}_i = \frac{f(\boldsymbol{y_i}) - f(\boldsymbol{y^{min}})}{f(\boldsymbol{y^{max}}) - f(\boldsymbol{y^{min}})},\ i = 1, \ldots, n \tag{3.15}$$

   The function chosen is then a normalized variation of a simple quadratic formulation with no cross-terms coefficient or linear component. More importantly this function satisfies all the characteristics described in section 3.2.1.

3. Noise is drawn from a normal distribution $\epsilon_i \sim N(0, \sigma),\ i = 1, \ldots, n$, where $\sigma = 0.1$

4. The cost values for the observations are then calculated as: $c_i = \tilde{c}_i + \epsilon_i,\ i = 1, \ldots, n$

In this first experiment we use five estimators on this DGP, all introduced in section 3.3: 1) SCKLS with coordinate-wise constraints, 2) SCKLS, 3) SCBLS with coordinate-wise constraints, 4) SCBLS with local Afriat constraints and 5) SCBLS with both coordinate-wise and local Afriat constraints. We detail the additional settings corresponding to each estimator. For the SCKLS estimators, the type of kernel used is

Gaussian, the grid is uniform and the number of grid points is approximately 100, specifically $m = 100$, that is 10 per axis for $d = 2$, $m = 64$ that is 4 per axis for $d = 3$ and $m = 81$ that is 3 per axis for $d = 4$. For the SCBLS, spline estimators, we use $r = 3$, with the number of bases $q = 6$, the knots being uniformly distributed. For SCBLS with the Afriat inequalities, the grid used is a uniform grid with 6 levels on each axis for $d = 2$, that is a total of 36 grid points, 5 levels on each axis for $d = 3$, that is a total of 125 grid points, and 4 levels on each axis for $d = 4$, that is a total of 256 grid points.

We start by showing the results corresponding to the fitting performance, represented by the Mean Squared Error (MSE) to the true function on the testing set. Using the DGP presented in Algorithm 1, we generate n observations for the training set that are used to determine the estimated function and n observations for the testing set that are used for performance measures. This MSE to the true function on the testing set corresponds to:

$$\text{MSE}_{\text{true-ts}} = 1/n \sum_{i=1}^{n} (\hat{c}_{ts_i} - \tilde{c}_{ts_i})^2 \tag{3.16}$$

where $\hat{c}_{ts_i}$ is the estimated cost value and $\tilde{c}_{ts_i}$ is the evaluation of the true cost function both at the outputs testing set levels $\boldsymbol{y_{ts_i}}$. The simulations are run on a computer with Intel i7-8700 GPU @ 3.20 GHz, 6 cores, and 16GB RAM on Matlab R2018b. The estimator SCBLS, derived from SCAM of Pya and Wood (2015) available in the R package “Scam", is recoded in Matlab.

The results are presented in Figure 3.1.

The results vary depending on the number of regressors. The SCBLS, spline estimators, have better performance than SCKLS for the two-regressor case, but as the regressors number is increased, SCKLS is more competitive. Additional constraints improve the out-of-sample fitting that we measure. However, not all constraints perform similarly. The SCBLS estimator with only local Afriat constraints on a set of grid points has the worse out-of-sample fitting performance among SCBLS estimators. The SCBLS estimator is very flexible and because unlike SCKLS, SCBLS does not use linear interpolation between grid points, it is possible the SCBLS estimator, with only local Afriat constraints, has a significant number of violations of the shape constraints and the local Afriat constraints are only local providing limited structure to the estimated function. This is confirmed when looking at the number of violations measure in the next paragraph.

Another metric that we use to evaluate the estimators is a function of shape constraints violations. The measure that we use for it is the percentage of violations of the Afriat inequalities at a set of testing points.

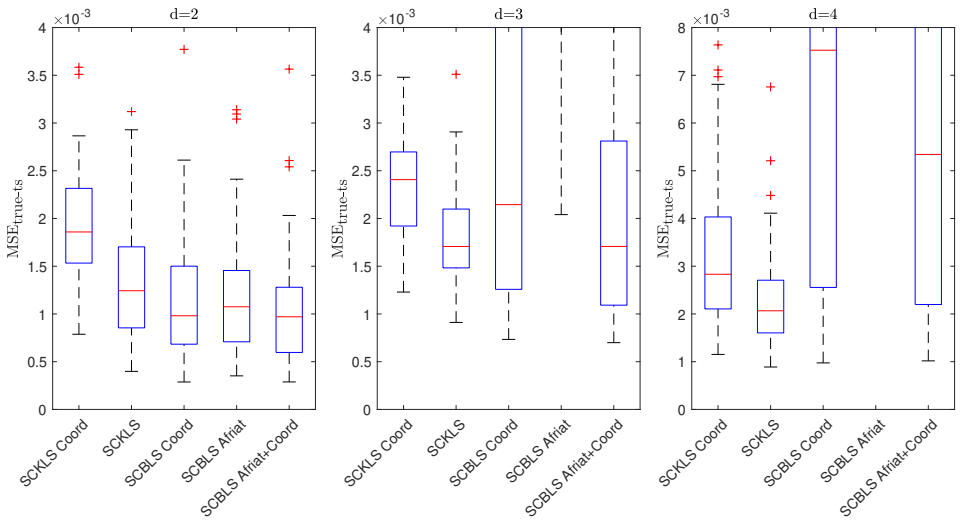

Figure 3.1: MSE in the base case with 100 observations and for a noise level with standard deviation $\sigma = 0.1$ on 50 simulations

This metric is introduced when the Afriat constraints are first mentioned in Section 3.2.3. We show the results for the ratio of violations of the Afriat inequalties concerning convexity in Figure 3.2.

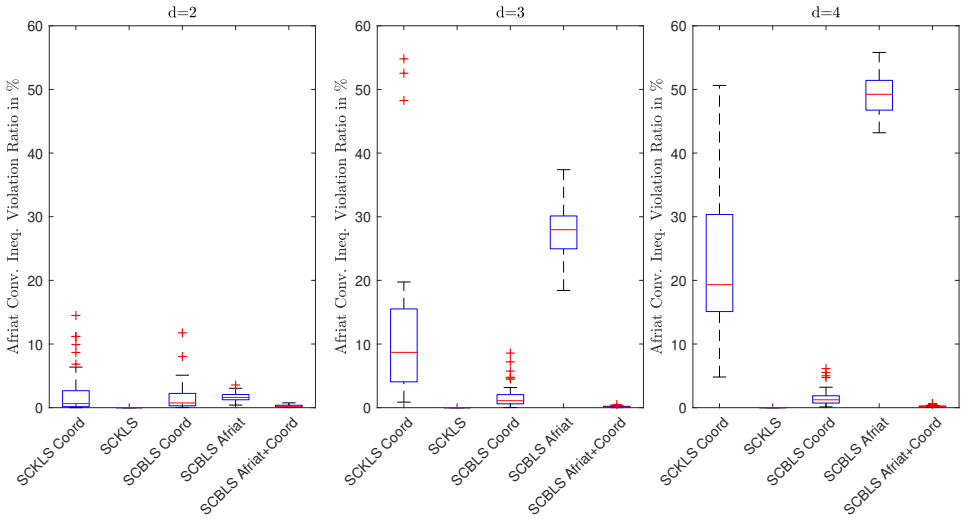

Figure 3.2: Percentage of Violations of the Afriat Inequalities concerning convexity on the set of testing points

The results confirm the shape constraints have a different impact when linear interpolation between grid points is not used. Local Afriat constraints are not enough for the SCBLS estimator. For SCBLS with Afriat constraints and coordinate constraints there are not many violations of convexity, but still fitting performance suffers as the number of regressors increases. This result is likely driven by some extreme values on the edges a topic we will return to in Section 3.4.2, Experiment 4.

3.4.1.2 Experiment 2 - Noise Levels variations relative to the Base case

In this experiment we follow the same DGP as the base case in Experiment 1 and vary the level of noise. We look at a case when noise is doubled, with the standard deviation used is $\sigma = 0.2$ versus $\sigma = 0.1$ in the base case and also the case when the noise is halved with $\sigma = 0.05$. We still use the out-of-sample fit as the performance metric defined in Experiment 1. Results are presented in Figure 3.3. Additionally, results concerning the violations of convexity on the set of testing points are presented in Figure 3.4. In both figures, each row corresponds to a different noise level σ and each column to a different number of regressors. Thus the middle row corresponds to the base case when $\sigma = 0.1$.

Figure 3.3 summarizes the fitting performance. The pattern across noise levels (from one row to another) is very similar. The ranking of performance of the estimators is not significantly affected by the noise level in contrast the number of regressors had a significant effect on the ranking of estimators. The conclusion is that the fitting performance is improved as the structure of the estimator increases, established in the base case, holds for other noise levels. Figure 3.4, summarizing shape constraint violations, shows differences in performance across estimators with variation in the noise level. In particular, when the number of regressors is small, the estimators with coordinate-wise constraints ('SCKLS Coord' and 'SCBLS Coord') show a significant increase in the number of convexity violations. For 'SCKLS Coord' estimator, $d = 2$, and $\sigma = 0.2$ the median percentage of violations is approximately 8% compared to approximately 2% in the base case; for $d = 3$, this difference is 20% in the noisy case compared to 10% in the base case. For the 'SCBLS Coord' estimator, the ratio increases are similar. For both estimators in the $d = 4$ case, the percentage of violations is already high in the base case, so the increase in noise does not impact the percentage of violations significantly.

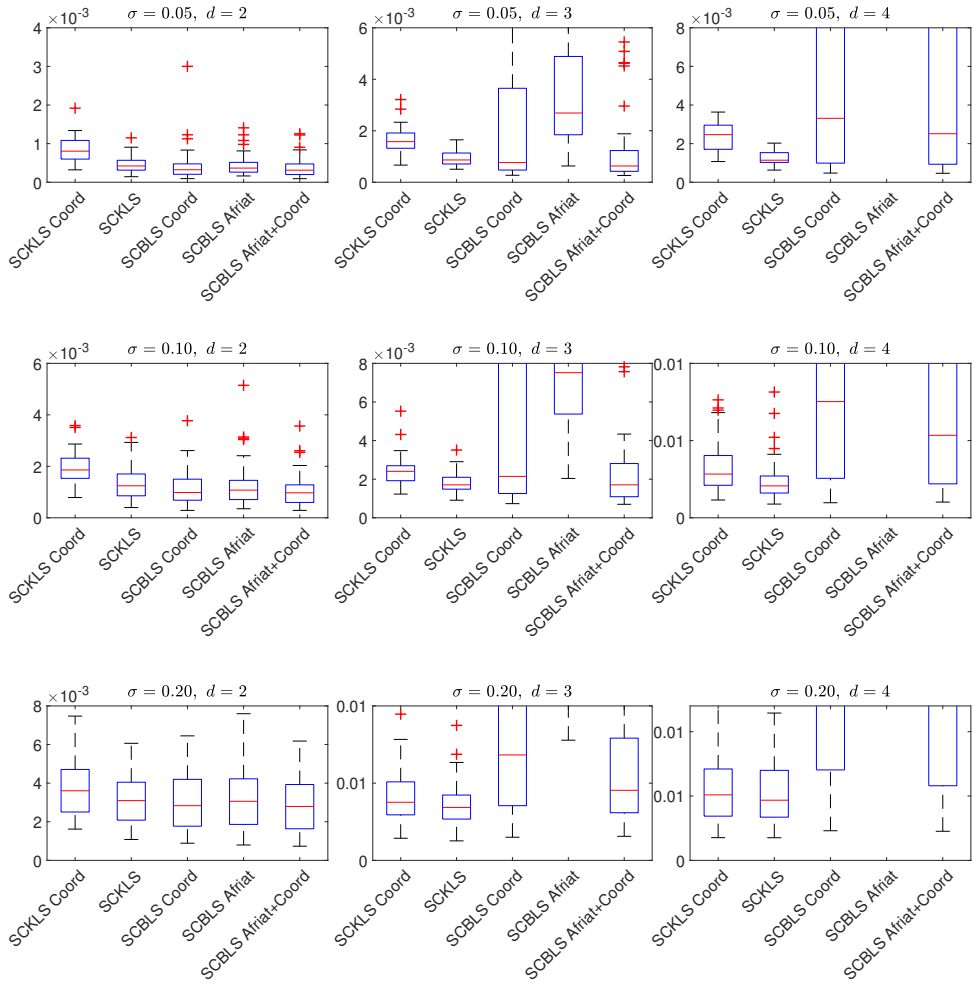

Figure 3.3: MSE measured relative to the true function for a set of 100 testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is 100 measured over 50 simulations.

3.4.1.3 Experiment 3 - Variation of the observation numbers relative to the Base case

In this experiment the number of observations varies from the base case 100 to 200 and 500. We also increase the number of testing points. The results for the MSE are presented in Figure 3.5 and the results with the percentage of violations of local Afriat inequalities concerning convexity are in Figure 3.6. The way both figures are built, each row corresponds to a different number of observations while each column to a different

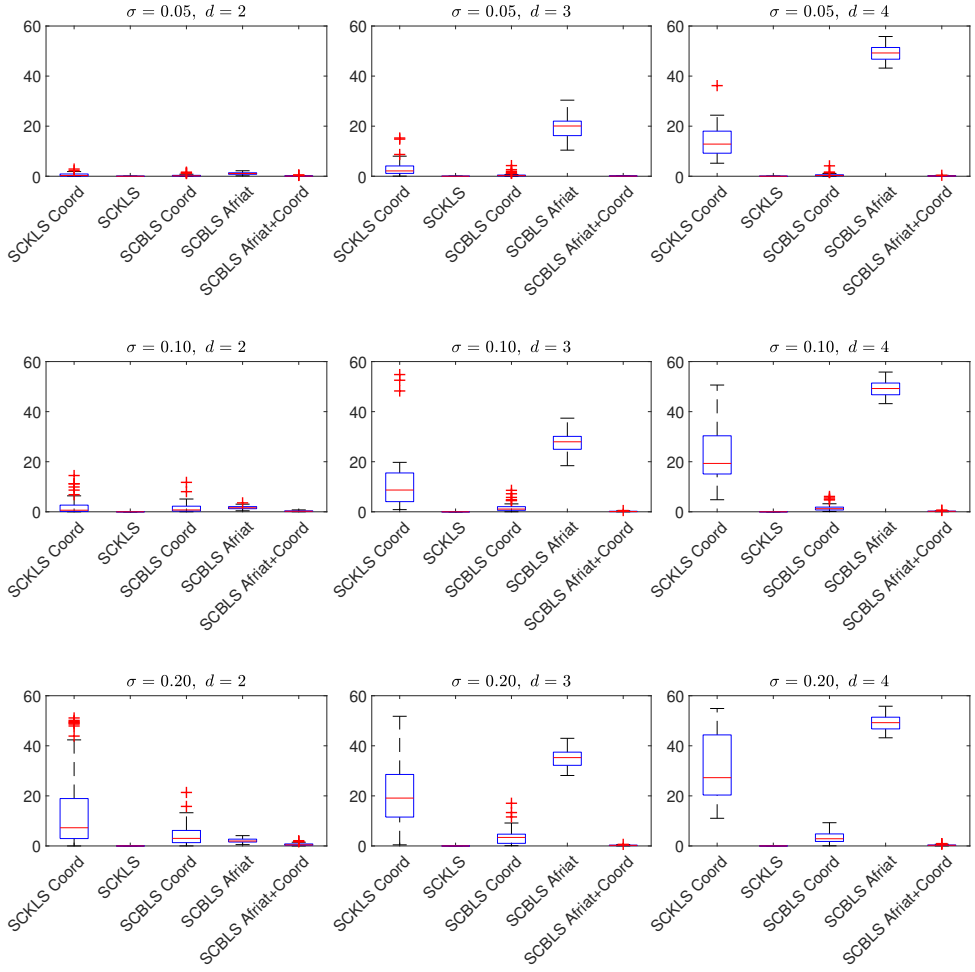

Figure 3.4: Percentage of violations of Afriat convexity for a set of 100 testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is 100 measured over 50 simulations.

number of regressors. In this case since the number of observations considered are $\{100,\ 200,\ 500\}$, the first row corresponds to the base case.

We first analyze the fitting performance results displayed in Figure 3.5. The results show that, as expected, increasing the number of observations results in smaller MSE values. For the two regressor case, $d = 2$, increasing the number of observations does not seem to change the ranking of the estimators. As

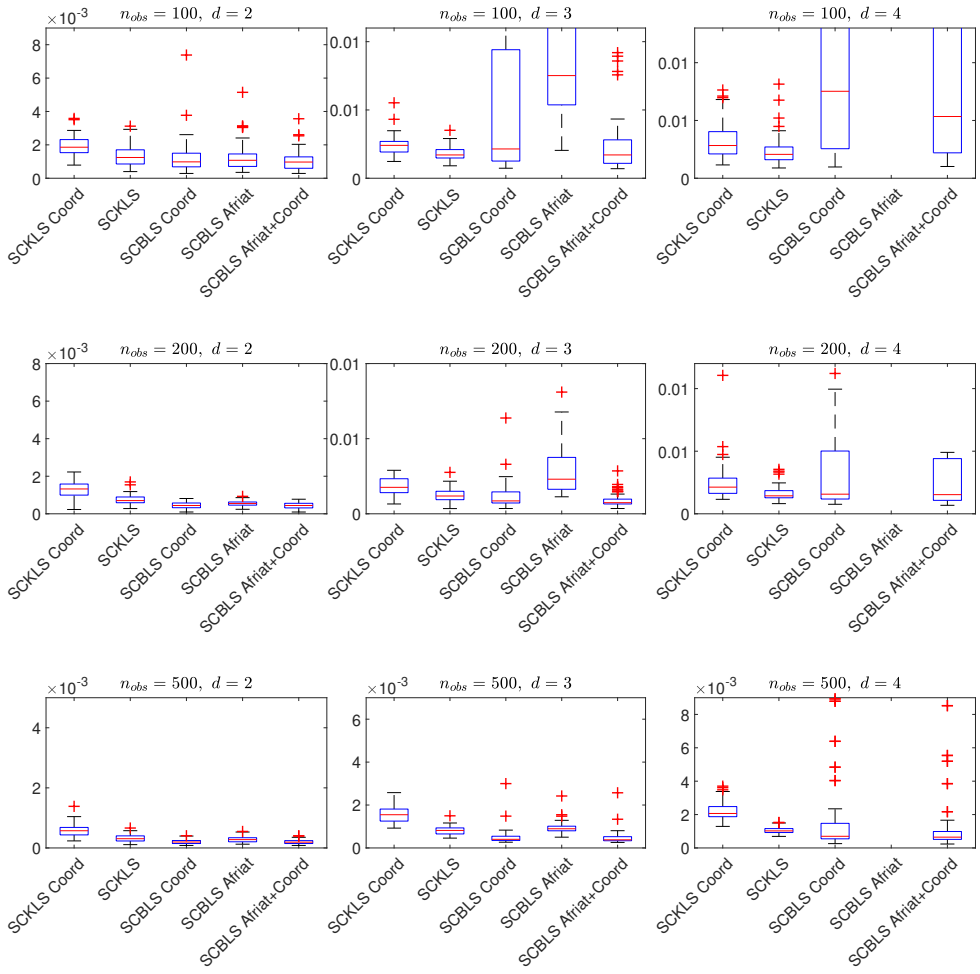

Figure 3.5: MSE measured relative to the true function for a set of $\{100,\ 200,\ 500\}$ testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is $\{100,\ 200,\ 500\}$ measured over 50 simulations.

it was for the base case, with 100 observations, the smooth estimators slightly outperform the piece-wise linear estimators. For $d = 3$ and $d = 4$, increasing the number of observations has a greater benefit for the smooth estimators. In particular, the SCBLS estimators with coordinate-wise constraints ('SCBLS Coord' and 'SCBLS Afriat+Coord') have better out-of-sample fitting performance than SCKLS estimators for 3 and 4 regressors when there are 200 observations and even more clearly with 500 observations. This result

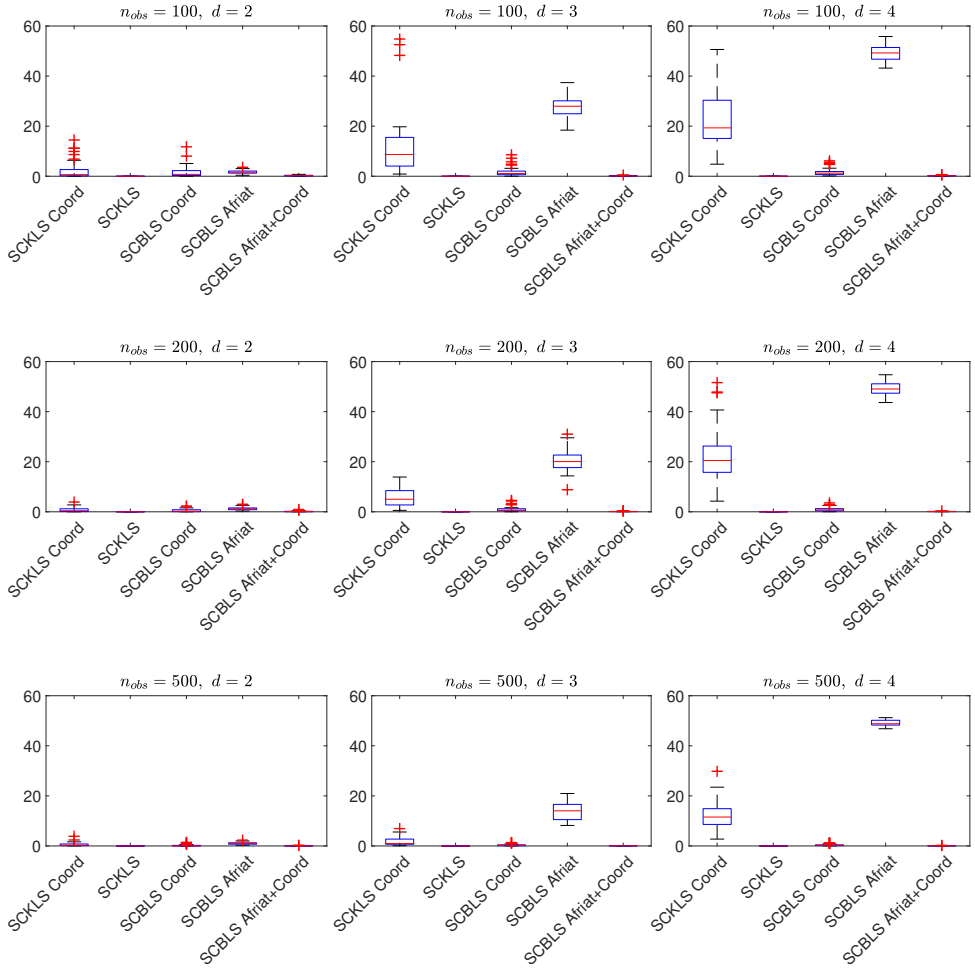

Figure 3.6: Percentage of violations of Afriat convexity for a set of $\{100,\ 200,\ 500\}$ testing points varying the noise standard deviation $\sigma = \{0.05, 0.1, 0.2\}$ and the number of observations is $\{100,\ 200,\ 500\}$ measured over 50 simulations.

corresponds to the performance of the estimators as measured by the median MSE values. Additionally, the SCBLS estimators have a high variance in performance compared to SCKLS. For the three regressor case, the variation in performance, as measured by the difference between the 25th and 75th percentiles, shows that the performance of the piece-wise linear estimators is more consistent. The smooth estimators outperform the piece-wise linear estimators based on median performance when the number of observations

is increased even in higher dimensions, but the overall dispersion is higher for smooth estimators relative to the piece-wise linear estimators.

The results for convexity violations are presented in Figure 3.6. The 'SCKLS Coord' estimator benefits the most from the increase in number of observations, in the $d = 3$ case, the percentage of convexity violations goes from 10% using 100 observations to almost none for 500 observations. The smooth estimator 'SCBLS Afriat', that only has the local Afriat constraints imposed on a set of control points, has a reduction from 30% to 15% in the same scenario. However, the 'SCBLS Afriat' estimator's fitting performance is still considerably worse than 'SCKLS Coord', as shown in Figure 3.5.

3.4.2 Additional estimators

3.4.2.1 Experiment 4 - Base case analysis with partial derivatives bounds for spline

For this experiment we add constraints to the SCBLS estimator to improve the boundary performance. We plotted the previous results and observed that poor fitting performance on out-of-sample data for the SCBLS estimator was coming from extreme variations in the estimated functions on the edges of the function's domain. To explore the boundary issue in higher dimensions, we impose bounds on the partial derivatives of the estimated functions at the set of control points. The formulation for each SCBLS is the same with the addition of the following constraints:

$$\text{PartDer}_\text{m} \leq b_{ik} \leq \text{PartDer}_\text{M}, \;\; i = 1, \ldots, m, \; k = 1, \ldots, d \tag{3.17}$$

where $\boldsymbol{b_i}$ are defined in Equation (3.12), PartDer_m and PartDer_M are respectively the lower bound and the upper bound for partial derivatives at the control points.

The value of the lower and upper bounds should be a function of parameters of the DGP and thus will vary across applications. For our DGP, we use $\text{PartDer}_\text{m} = 0.01$ and $\text{PartDer}_\text{M} = 1$. We present the results obtained for the SCKLS estimators without bounds and the three SCBLS versions with the bounds, the DGP is the base case introduced in Experiment 1. In Figure 3.7, the out-of-sample fitting performance is reported and in Figure 3.8 the Afriat convexity violations are reported.

The results show that adding these constraints significantly improves the SCBLS estimators performance. The median MSEs are almost halved in both $d = 3$ and $d = 4$ cases for the SCBLS estimators with

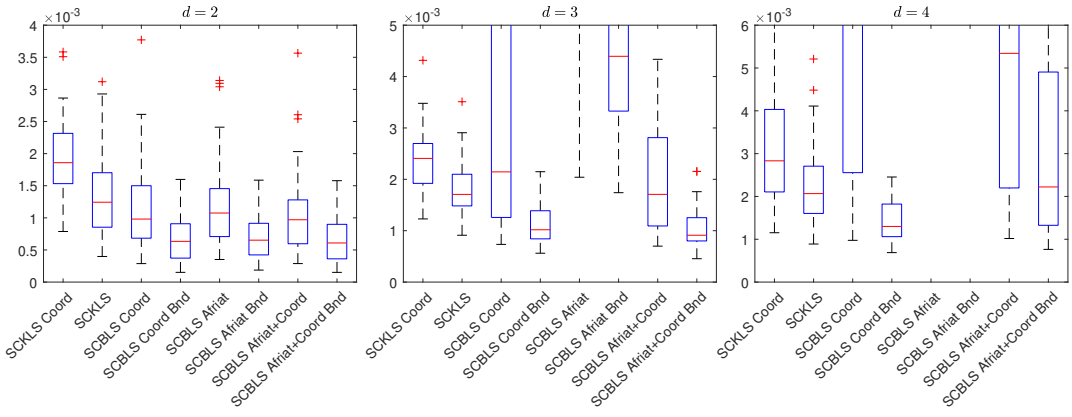

Figure 3.7: MSE measured relative to the true function for a set of 100 testing points with the noise standard deviation $\sigma = 0.1$ and the number of observations is 100 measured over 50 simulations adding results from estimator with slope bounds.

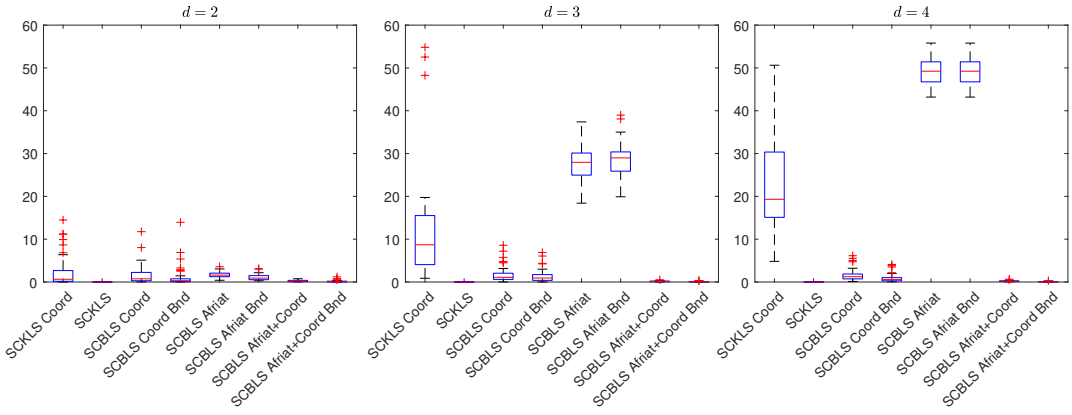

Figure 3.8: Percentage of violations of Afriat convexity for a set of 100 testing points with the noise standard deviation $\sigma = 0.1$ and the number of observations is 100 measured over 50 simulations adding results from estimator with slope bounds.

coordinate-wise constraints ('SCBLS Coord' and 'SCBLS Afriat+Coord') and imposing these bounds. Now the SCBLS estimators with the bounds outperform SCKLS. Bounding the slopes appears to control the variation of the SCBLS class of estimators near the edge of the function domain improving their out-of-sample fitting performance.

3.4.3 Summary and recommendations

The results from the simulations can be summarized as follows:

1. **Piece-wise linear approximation**: Using the piece-wise linear approximation versus smooth estimators is an advantage because the linear interpolation between grid points enforces the convexity on the entire domain of the function. However smooth estimators with shape constraints outperform the piece-wise linear estimators in low dimensional cases or when the training set is large and the noise is not too large. With a large number of regressors and few data points, the estimators using the piece-wise linear approximations are preferred and provide consistent results.

2. **Coordinate-wise constraints approximation**: First the loss in fitting performance between 'SCKLS Coord' and 'SCKLS' shows that there is significant benefit to imposing the Afriat version of the shape constraints. Second, the constraints used in 'SCBLS Coord' are stronger than the coordinate-wise constraints, as they are sufficient but not necessary conditions of them. They also apply on the entire domain of definition which appears to be be a big advantage compared to constraints only imposed locally on a set of control points.

3. **Local Afriat Inequalities as constraints**: Using only local Afriat inequalities on a smooth estimator performs well for cases with two regressors, but as the number of regressors increases the number of shape constraints violations increase. A very large number of observations are required to obtain comparatively good estimates. However, compared to using the Pya and Wood (2015) coordinate-wise sufficient constraints on a smooth estimator, also including the Local Afriat Inequalities provides a small boost in performance.

Thus Table 3.2 can be inferred to report the recommended estimator for different cases:

3.5 Application to the cost function estimation of the US hospital sector

In this section, we present the results of the splines-based shape-constrained non-parametric estimator applied to a dataset containing the number of procedures performed and the cost for a sample of US hospitals. Based on the results of the simulations and the conclusions presented in Table 3.2, 'SCBLS Afriat+Coord' is selected for this data set that has slightly more than 400 observations per year and four regressors. Thus the following section will focus on the analysis using 'SCBLS Afriat+Coord' as the estimator of the cost

Table 3.2: Optimal estimators in a case of non-extreme noise levels based on our simulation results depending of data sets characteristics

		Number of Regressors	
		2	> 2
Number of Observations	< 400	SCBLS ('Afriat+Coord')	SCKLS
	> 400	SCBLS ('Afriat+Coord')	SCBLS ('Afriat+Coord') / SCKLS (very noisy data)

function. More results comparing alternative estimators and results of other estimators applied to the hospital data are available in Appendix B.2.

3.5.1 Presentation of the data set

The hospital dataset joins cost data from the American Hospital Association's (AHA) Annual Survey Databases and outputs data from the Healthcare Cost and Utilization Project (HCUP) National Inpatient Sample (NIS) core file. We focus on years 2007 and 2008. The AHA Annual Survey contains an extensive list of of metrics for every US hospital. The cost measure built takes into consideration payroll, employee benefits, depreciation, interest, supply expenses and other expenses. If we could not map the AHA data to the outputs measured in the HCUP data, we excluded that particular AHA data record from the analysis. Among the output data, there are several observations which are not similar to other hospitals in the sample. To allow to focus on trends among the most common sizes of hospitals without being effected by these potential outliers, all records with more than the 90th percentile in an output category were removed from the analysis. The trimmed version of the data that is used in the application is summarized in Table 3.3.

3.5.2 Results

3.5.2.1 Analysis on the marginal rates of substitution

We perform an analysis on the marginal rates of substitution (MRS) concerning the outputs of the industry. The levels for minor and major diagnostic procedures are held at their median level and we consider the marginal rate of substitution between the major therapeutic procedures and minor therapeutic procedures, as MRS_{MT-mT}, it can be expressed as:

Table 3.3: Data Summary

	2007 (437 observations)				
	MinDiag	MinTher	MajDiag	MajTher	Cost($)
Mean	1505	3798	84	2106	80M
Skewness	2	1	1	1	1
25th Percentile	82	407	6	200	20M
50th Percentile	554	1869	43	1097	54M
75th Percentile	2182	6002	132	3177	112M
	2008 (421 observations)				
	MinDiag	MinTher	MajDiag	MajTher	Cost($)
Mean	1563	3935	92	2274	88M
Skewness	2	1	1	2	2
25th Percentile	80	405	7	215	22M
50th Percentile	607	2155	47	1193	59M
75th Percentile	2340	6136	142	3425	118M

$$\text{MRS}_{MT-mT} = \text{MC}_{MT}/\text{MC}_{mT} \tag{3.18}$$

where MC_{MT} and MC_{mT} are the marginal costs of the major and minor therapeutic procedures respectively. These marginal costs are evaluated for combinations of outputs at different percentiles of major and minor therapeutic procedures. They are evaluated as numerical approximations of the derivatives of the cost function at these levels. The detailed results for the MCs are detailed in the next paragraph, first the MRSs are reported in Figure 3.9.

To understand the interpretation of the MRS_{MT-mT} values consider the calculation $\text{MRS}_{MT-mT} = 2$, this calculation implies that conditional on the current number of procedures and maintaining the same cost level, minor therapeutic procedures should be reduced by 2 in order to perform one more major therapeutic procedure. Two main insights can be inferred from these results; the trends are consistent across both years, and the scale being lower in 2008. First, for hospitals that already perform many minor therapeutic procedures, the MRS are smaller indicating the substitution of minor to major therapeutic procedures requires a smaller reduction in minor therapeutic procedures to increase major therapeutic procedures compared to hospitals that perform few minor therapeutic procedures. A possible explanation is economies of scale ex-

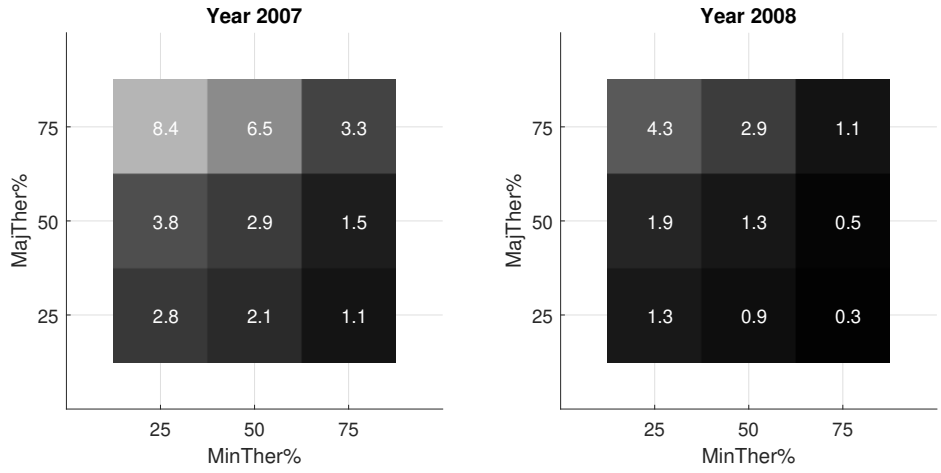

Figure 3.9: MRS between Major Therapeutic procedures and Minor Therapeutic procedures obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values.

ist. Specifically, large hospitals will have more resources over which they can spread minor therapeutic procedures while changing some resources over to focus on major therapeutic procedures. The second insight corresponds to the fact that the MRS is higher for higher percentiles of major therapeutic procedures. One possible explanation is that hospitals have more specialized equipment and that increasing the number of major therapeutic procedures by reducing the number of minor therapeutic procedures becomes more difficult.

Eventually it depends how much each type of procedures is valued and if the hospital can influence patient mix but these results indicate that a hospital with a lot of minor therapeutic procedures would benefit to diversify with more major therapeutic procedures as the MRS is close to 1 and even smaller than 1 for 2008.

3.5.2.2 Marginal cost analysis

The marginal costs (MC) that are used for the computation of the marginal rates of substitution (MRS) above are reported in details in Figure 3.10 for the Major Therapeutic procedures and Figure 3.11 for the Minor Therapeutic procedures respectively. The values are in k$ per procedure.

The marginal costs can be interpreted as follows. If we consider the marginal cost of Major Therapeutic

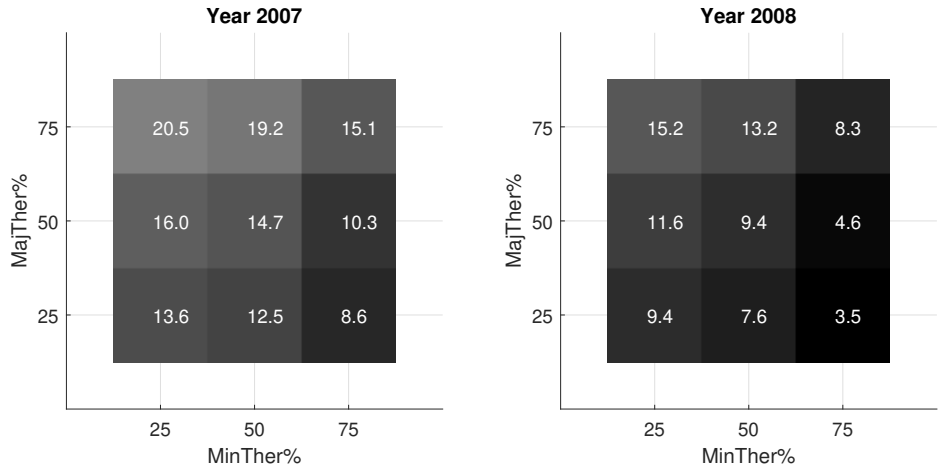

Figure 3.10: MC of the Major Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values.

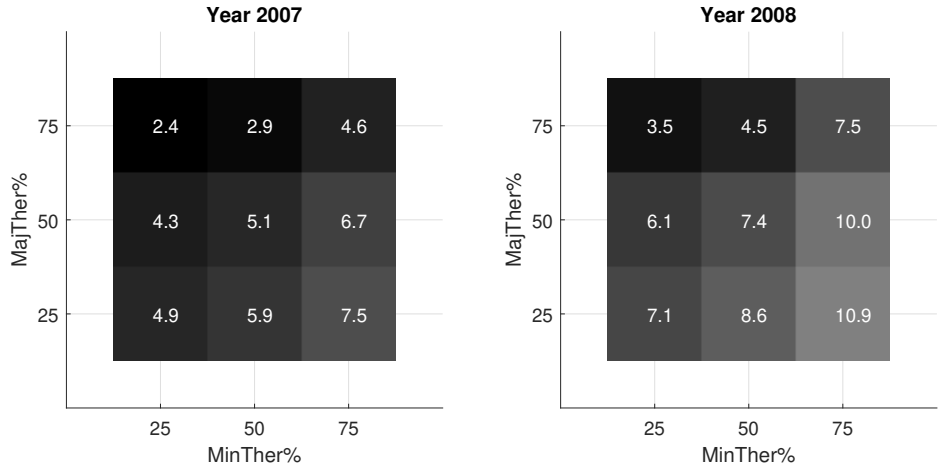

Figure 3.11: MC of the Minor Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCBLS Afriat+Coord' for years 2007 and 2008, evaluated at different percentile values.

procedures, $\text{MC}_{MT} = 10$ k$ per procedure, this value means that conditional on the current output and cost levels, the cost of performing an additional Major Therapeutic procedure is 10 k$. Therefore, marginal cost is a function of the current output and cost levels. Two main results are observed. First, looking

at Figure 3.10, the marginal cost of Major Therapeutic procedures increases with the number of Major Therapeutic procedures performed. For instance in year 2007, for a fixed percentile of 50th for Minor Therapeutic procedures, the marginal cost of Major Therapeutic procedures is 14.7 k$ per procedure for a level of Major Therapeutic procedures at the 50th percentile and 19.2 k$ per procedure for a level at the 75th percentile. Increasing marginal cost is commonly caused by organizational issues when hospitals are performing a lot of a particular type of procedures and additional increases in procedures of a particular procedure will require scarce resources. For example a particular piece of equipment or structural resource is then saturated, increasing costs. Second, the marginal cost of a Major Therapeutic decreases when the number of Minor Therapeutic procedures performed increases. For instance for a fixed percentile of 50th for Major Therapeutic procedures, the marginal cost of the Major Therapeutic procedures is 14.7 k$ per procedure for a level of Minor Therapeutic procedures at the 50th percentile and 10.3 k$ per procedure for a level at the 75th percentile. This actually corresponds to what was observed with the analysis of the marginal rates of substitution, the benefits for a hospital to diversify and produce more Major Therapeutic procedures are positive if the hospital already performs a large number of Minor Therapeutic procedures. We refer to the previous paragraph for more details about economies of scale argument which is the basis of this phenomenon. Both results are also observed for Minor Therapeutic procedures in Figure 3.11, and the same reasoning can be applied to this class of procedures.

3.6 Conclusion

This chapter provided an analysis of shape constraints approximations commonly used in shape-constrained non-parametric estimators, a type of estimators that we argue has the advantages to be more flexible than fully parametric approaches, while maintaining the desired level of interpretability to obtain valuable insights about the industry considered. The two approximations analyzed were piece-wise linear function versus smooth function and global constraints versus coordinate-wise constraints. B-spline estimators with shape constraints were developed for the smooth cases and Local Linear Kernel estimators with shape constraints were used for the piece-wise linear cases. The estimators were tested on a set of simulations to cover a wide range of settings, the criteria of evaluation being the fitting performance and the violations of the shape constraints. The results of the simulations are that for noisy data sets and for which you have more than two regressors and less than 400 observations, the piece-wise linear estimator is preferred be-

cause global shape constraints can be imposed on the entire domain of the function. For other scenarios the smooth B-spline estimator is provides slightly better performance, even if violations of the shape constraints increases slightly. The implementation of one of the spline estimators developed on a cost function of the US hospital industry to analyze marginal rates of substitution indicated hospitals performing many minor therapeutic procedures and few major therapeutic procedures would benefit from re-balancing their outputs production to increase their productivity where possible. Future research could be done in the application part to extend the hospital analysis or in the methodology part to optimize the selection of the settings for the estimators, such as the number of knots and their distributions for the B-spline estimators.

4. CONCLUSIONS

Global interest in production economics has increased due to the rapid development of data analytics and its applications. Estimation methods also need to evolve, and cost function estimation offers significant potential improvement through the use of shape-constrained non-parametric estimators. Two main axes of research were considered and led us to determine recommendations for users interested in such approaches.

First, we discussed the importance of direction selection for estimators with stochastic directional distance functions. The CNLS-d estimator is a shape-constrained non-parametric estimator based on the stochastic directional distance function. Using this estimator and Monte Carlo simulations, we provided guidelines for selecting a well performing direction, to obtain a good fitting performance to the true function: Guideline 1: If the user has an idea of the noise direction, assign this direction as the direction used in the estimator; Guideline 2: Select a direction orthogonal to the data. Since knowing the noise direction is not very common, most users rely on the orthogonal direction. The CNLS-d estimator with the direction orthogonal to the data was implemented for estimating a cost function in the US hospitals industry. The results were compared to the results of a quadratic estimator, that is fully parametric and a linear kernel regression, that is fully non-parametric. The CNLS-d with direction selection is more flexible to avoid possible functional form misspecification coming from a fully parametric model and it has more structure than the fully non-parametric model to help with the estimation and interpretation of valuable metrics, for instance marginal costs. The main insight gained for the US hospitals industry is that the Most Productive Scale Sizes are small, which indicates that clusters of smaller hospitals would be more productive than large hospitals.

Second, we analyzed two shape approximations commonly used for shape-constrained non-parametric estimators: piece-wise linear versus smooth functional forms, and global constraints versus coordinate-wise shape constraints. We developed an estimator based on B-splines, called SCBLS, to become the base of the smooth estimators used for the comparison. The piece-wise linear estimators were based on a shape-constrained local linear kernel regression, called the SCKLS estimator. The results of an exhaustive range of experiments, led us to conclude that SCKLS is the best estimator for problems with both small datasets (less than 400 observations) and more than two regressors, and that SCBLS is the best estimator for other

problems, even if the shape constraints are not enforced globally (over the entire domain of the function). SCKLS still performs well in all cases. The standard deviations for its fitting performances were smaller, particularly for some of the larger datasets with four regressors. The result suggests it might be a preferred estimator for cases with high noise levels. We used SCBLS to estimate the US hospital industry cost function. Our study of the marginal rates of substitution determined that hospitals that perform many minor therapeutic procedures would be more productive if they re-balance their outputs to perform more major therapeutic procedures, if customer in-take practices allow.

In summary, based on our analytical results, we suggested recommendations for users of non-parametric shape-constrained estimators. First, we proposed a method to select the direction for an estimator using a stochastic directional distance function. Second, we provided information about the performance of different approximations to shape constraints of non-parametric estimators to help with estimator selection. We developed several estimators for the latest analysis including the B-spline based estimators, SCBLS, for which promising results were obtained. From the applications subsections, we shared insightful results on the US hospitals industry.

Future research could build on the results discussed in Chapter 2 by continuing studying the possibility of developing an estimator that would allow using several directions depending on the data and still maintaining the shape constraints. Also more study of the SCBLS estimator could reveal the benefits of tailoring the SCBLS estimator to specific applications.

REFERENCES

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

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

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

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

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

S. Atkinson, C. Cornwell, and O. Honerkamp. Measuring and decomposing productivity change: stochastic distance function estimation versus data envelopment analysis. *Journal of Business & Economic Statistics*, 21(2):284–294, 2003. ISSN 0735-0015.

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

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

N. Bloom and J. Van Reenen. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408, 2007.

N. Bloom and J. Van Reenen. Why do management practices differ across firms and countries? *Journal of Economic Perspectives*, 24(1):203–24, 2010.

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

R. Carroll, D. Ruppert, L. Stefanski, and C. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Boca Raton, FL, 2006. ISBN 9781420010138.

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

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

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

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

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

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

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

C. Daraio and L. Simar. Efficiency and benchmarking with directional distances: a data-driven approach. *Journal of the Operational Research Society*, 67(7):928–944, 2016.

C. De Boor, C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.

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

P. Du, C. F. Parmeter, and J. S. Racine. Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, pages 1347–1371, 2013.

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

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

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

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

R. Färe, C. Pasurka, and M. Vardanyan. On endogenizing direction vectors in parametric directional distance function-based models. *European Journal of Operational Research*, 262(1):361–369, 2017.

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

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

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

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

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

L. A. Hannah and D. B. Dunson. Multivariate convex regression with adaptive partitioning. *The Journal of Machine Learning Research*, 14(1):3261–3294, 2013.

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

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

B. Hollingsworth. Non-parametric and parametric applications measuring efficiency in health care. *Health care Management Science*, 6(4):203–218, 2003.

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

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

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

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

T. Kuosmanen. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the stoned method in the finnish regulatory model. *Energy Economics*, 34(6):2189–2199, 2012.

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

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

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

T. Kuosmanen, A. Johnson, and A. Saastamoinen. Stochastic nonparametric approach to efficiency analysis: A unified framework. In *Data envelopment analysis*, pages 191–244. Springer, 2015.

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

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

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

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

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

C. K. Lovell, P. Travers, S. Richardson, and L. Wood. Resources and functionings: a new view of inequality in Australia. In *Eichhorn (Ed.), Models and Measurement of Welfare and Inequality*, pages 787–807. Springer, Berlin, Germany, 1994.

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

C. Manski. *Partial identification of probability distributions*. Springer Series in Statistics. Springer, New York, NY, 2003. ISBN 9780387004549.

G. S. Olley and A. Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996. ISSN 00129682, 14680262.

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

N. Pya and S. N. Wood. Shape constrained additive models. *Statistics and Computing*, 25(3):543–559, 2015.

J. Racine and Q. Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130, 2004.

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

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

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

R. C. Sickles, D. H. Good, and L. Getachew. Specification of distance functions using semi-and nonparametric methods with an application to the dynamic performance of eastern and western european air carriers. *Journal of Productivity Analysis*, 17(1-2):133–155, 2002.

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

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

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

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

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

X.-g. Zhao, G.-w. Jiang, D. Nie, and H. Chen. How to improve the market efficiency of carbon trading: A perspective of china. *Renewable and Sustainable Energy Reviews*, 59:1229–1245, 2016.

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

APPENDIX A

APPENDIX OF CHAPTER 2

The appendix is composed of the following parts:

- Properties of Directional Distance Functions and CNLS-d (A.1)
- Monte Carlo, Additional Experiments (A.2)
- Detailed results for the Hospital Application (A.3)

A.1 Properties of Directional Distance Functions and CNLS-d

A.1.1 Direction Selection in Directional Distance Functions

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

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

Proof. Rewrite Problem (2.10) from Section 3.2 as

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

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

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

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

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

□

A.1.2 Details of CNLS-d

An alternative expression for CNLS-d (cf. equations (16)-(16c) from Section 5.1) is given by:

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

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

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

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

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

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

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

$$\alpha_i = -\boldsymbol{\beta_i'}\,\boldsymbol{x_i} + \boldsymbol{\gamma_i'}\,\boldsymbol{y_i} + \epsilon_i, \text{ for } i = 1, \ldots, n. \tag{A.5}$$

A.1.3 Different Directions for Different Groups in CNLS-d

Consider the case where all observations have the same input level and produce two outputs and estimate the isoquant. Define two groups of observations G_1 and G_2 such that $|G_1 \cup G_2| = n$ and $G_1 \cap G_2 = \emptyset$.[1] Using the notation in A.1.1, the direction vector for the first group of observations G_1 is $\boldsymbol{g}^{y_{G_1}}$ and it's $\boldsymbol{g}^{y_{G_2}}$ for the second group of observations G_2.

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

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

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

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

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

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

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

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

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

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

where

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

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

Proof. Consider the Afriat inequalities in the context of cost isoquant estimation. One of the conditions of Equation (2.16) is:

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

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

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

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

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

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

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

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

which, after simplifying, becomes:

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

□

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

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

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

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

where

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

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

Proof. Reverse the inequality sign in Equation (A.8):

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

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

□

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

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

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

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

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

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

two groups. This does not affect the proof of the condition.

□

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

A.1.3.1 Simulations to investigate the frequency with which multiple directions leads to violations

We run simulations to investigate the effects of using multiple directions. We use the same DGP as stated in Section 2.5, Example 1. However, we define two groups and assign different directions for each one of them:

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

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

and,

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

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

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

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

significant number of violations of the maintained assumptions.

A.2 Additional Experiments

A.2.1 Experiments Related to Section 2.4.1 - with the Linear Estimator.

A.2.1.1 Measuring MSE Example, Section 2.4.1 - Noise Generated in a Common and Prespecified Direction θ_f

This section describes the simulations and the results for the fixed noise direction case referenced in Section 2.4.1.

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

1. The output, $\tilde{y}_i$, is drawn from the continuous uniform distribution $U\,[0, 1]$
2. The cost is calculated as $\tilde{c}_i = \beta_0\, \tilde{y}_i$, where $\beta_0 = 1$.
3. In the case of fixed direction, the noise term is determined as:
   (a) l_{ϵ_i} is the scalar length that is drawn from a normal distribution, $N\,(0, \lambda\, \epsilon_0)$, λ is prespecified and an initial value for the standard deviation, ϵ_0, is calculated as in Equation (2.11) in Section 2.4.1.:

$$\epsilon_0 = \frac{1}{2}\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{y}_i - \bar{y}\right)^2} + \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\tilde{c}_i - \bar{c}\right)^2}\right], \tag{A.18}$$

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

$$\begin{pmatrix} y_i \\ c_i \end{pmatrix} = \begin{pmatrix} \tilde{y}_i \\ \tilde{c}_i \end{pmatrix} + \begin{pmatrix} \epsilon_{y_i} \\ \epsilon_{c_i} \end{pmatrix},\ i = 1, \ldots, n. \tag{A.19}$$

Apply the DGP described above to generate a training set, $(y_{tr_i}, c_{tr_i})\,,\ i = 1, \ldots, n_{tr}$, and a testing set

Figure A.1: Linear function data generation process with fixed noise direction

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

We test the set of directions corresponding to the angle $\theta_t \in \{0, \pi/8, \pi/4, 3\pi/8, \pi/2\}$. If the direction of the noise, θ_f, matches the direction used in the DDF, θ_t, then the smallest MSE results for all cases.

A.2.1.2 Results: Fixed Noise Direction

Table A.1 reports the MSE computed by comparing the estimated function to the true function and Table A.2 reports the MSE computed by comparing the estimated function to the testing set.

In Table A.1, the direction for the DDF corresponding to the smallest MSE always matches the noise direction in the DGP. Further for more than 70% of the cases tested there is more than a 50% decrease in MSE by using the correctly specified direction compared to the next best direction tested, which was not as large in the random direction case in Table 2.1 of Section 2.4.1. In other words, when endogeneity is severe, the benefits of using a DDF with a well-selected direction are potentially large.

Table A.2 is consistent with the results observed in the random noise case, in Table 2.2 of Section 2.4.1. The DDF directions corresponding to the smallest MSE values are those matching the directions used for the MSE computation. Thus, the proposed radial MSE measure addresses the challenge of measuring performance in applications with a testing dataset.

A.2.1.3 Monte Carlo Simulations - Experiments, Section 2.5.2.3 - Experiment 3. Base case with fixed noise direction and different noise levels

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

A.2.2 Experiments related to Section 2.5.2 - with CNLS-d

Here we complete Section 2.5.2 with additional experiments and we follow the numbering experiments numbering established then.

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

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

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

A.2.2.1 Experiment 5: Base case with different distributions for the initial observations on the true function

In Experiment 5, we extend the analysis performed in Experiment 4. We consider additional distributions of the DGP for the angle, $\theta_i,\ i = 1, \ldots, n$ and see how it affects the optimal direction. Unlike

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

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

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

Experiment 4, we don't consider only normal distributions, instead we consider the following: a normal distribution, $N\left(\frac{\pi}{4}, \frac{\pi}{16}\right)$, and two gamma distributions, $\Gamma\left(3, \frac{\pi}{2}\right)$ and $\Gamma\left(.5, \frac{\pi}{24}\right)$. For the gamma distributions, the first parameter corresponds to the shape coefficient and the second the scale coefficient. Each distribution is later referenced respectively as $Normal$, $Gamma_1$ and $Gamma_2$. We truncate the tails of the

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

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

Note: Displayed are measured values multiplied by 10^4

distribution so that the generated angles fall within the range $[0, \pi/2]$. Noise is specified as in Experiment 1. In Figure A.2, the distributions of the angles θ_i are illustrated and in particular the median values are highlighted. Table A.4 reports the results of this experiment.

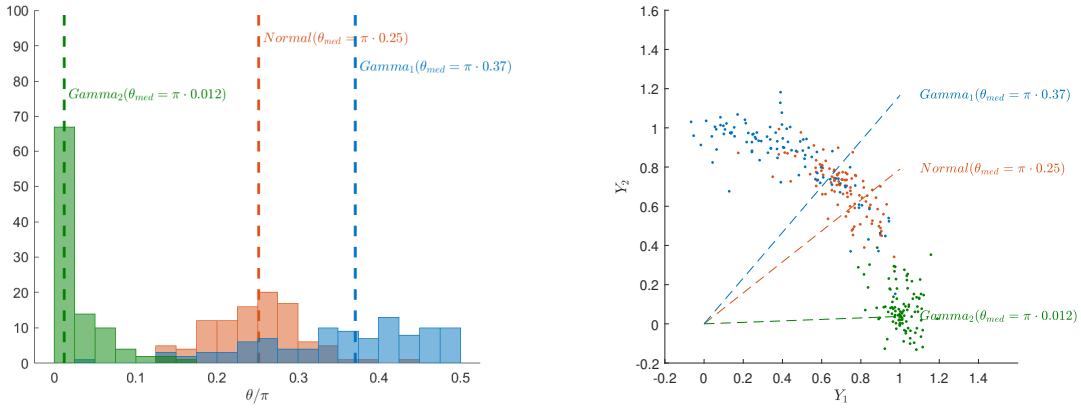

Figure A.2: (a) On the left, histogram of the angles θ_i and its median obtained for each distribution when running a simulation with 100 observations. (b) On the right, the corresponding observations and the median of the angles θ_i for each distribution for the same simulation as the histogram (a).

Two main conclusion can be drawn from the results in Table A.4. First, the smaller the variance of the data distribution, the greater is the importance of direction selection. Looking at the differences between

Table A.4: Experiment 5: Values of radial MSE relative to the true function varying the CNLS-d direction and type of direction used for the DGP.

	CNLS-d Direction Angle				
Distribution	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
$Normal$	8.45	3.04	**1.96**	3.01	8.60
$Gamma_1$	29.34	6.92	3.27	**2.54**	3.39
$Gamma_2$	**6.62**	9.69	19.19	72.55	598.97
Note: Displayed are measured values multiplied by 10^4.					

the two gamma distributions, $Gamma_1$ has a larger tail than $Gamma_2$, which means the observations for $Gamma_2$ have a smaller variance. Table A.4 indicates that the MSE increases rapidly with deviations from the optimal direction when variance of observations is smaller as with $Gamma_2$ compared to $Gamma_1$. Second, among the directions tested, θ_i, MSE is minimized for the direction closest to the direction corresponding to the median of the distribution. This second point supports the selection approach proposed in Section 2.6.

A.2.2.2 Experiment 6: Adaptation of the Base Case to a 3-Dimensional Case

We adapt the DGP from Experiment 1, the base case. We consider a fixed input level and approximate a three output isoquant, $Q = 3$. Indexing the outputs by q and observations by i, we define the outputs,

$$y_{qi} = \tilde{y}_{qi} + \epsilon_{qi},\ q = 1, \ldots, Q,\ i = 1, \ldots, n, \tag{A.20}$$

where $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ is the observation on the isoquant and $\boldsymbol{\epsilon}_{\boldsymbol{i}}$ is the noise. The output levels $\tilde{y}_{qi},\ q = 1, \ldots, Q,\ i = 1, \ldots, n$ are generated:

$$\tilde{\boldsymbol{y}}_i = \frac{\boldsymbol{l}_i}{||\boldsymbol{l}_i||_2},\ i = 1, \ldots, n \tag{A.21}$$

where $l_{qi},\ q = 1, \ldots, Q,\ i = 1, \ldots, n$, are drawn randomly from a continuous uniform distribution, $U\,[0, 1]$.

The noise terms $\boldsymbol{\epsilon}_i,\ i = 1, \ldots, n$ is adapted to the 3-dimensional isoquant:

$$\boldsymbol{\epsilon}_i = l_{\epsilon_i}\ \boldsymbol{v}_i,\ i = 1, \ldots, n \tag{A.22}$$

where the length l_{ϵ_i} is drawn from the normal distribution $N\,(0, \lambda)$, and $v_{qi} = \frac{v^*_{qi}}{||\boldsymbol{v}_i||_2},\ q = 1, \ldots, Q,\ i = 1, \ldots, n$ for which v^*_{qi} are drawn from a continuous uniform distribution $U\,[-1, 1]$.

In Experiment 6, 19 directions are considered for the CNLS-d estimators. The directions are determined using the following steps:

1. enumerate all 3 component vectors, corresponding to $\mathbb{R}^3$ with elements from the set $\{0, 0.5, 1\}$ and excluding $(0, 0, 0)$;
2. normalize the direction vectors dividing them by their respective Euclidean norms;
3. eliminate duplicates

The 19 directions are represented by the markers in Figure A.3 and create a balanced grid on the eighth of a unit sphere, our isoquant. The median direction is $[1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}] = [.58, 58, .58]$. The standard

deviation of the normal distribution is $\lambda = 0.1$. We perform this experiment 100 times for each direction. We report the averaged radial MSE values on a testing set of n observations lying on the true function in Table A.5. In addition to the table, the MSE results are also illustrated in Figure A.3 where the size of the markers has a positive affine relation with the MSE values and that in the color range from yellow to red, with larger the MSE values associated with more red markers.

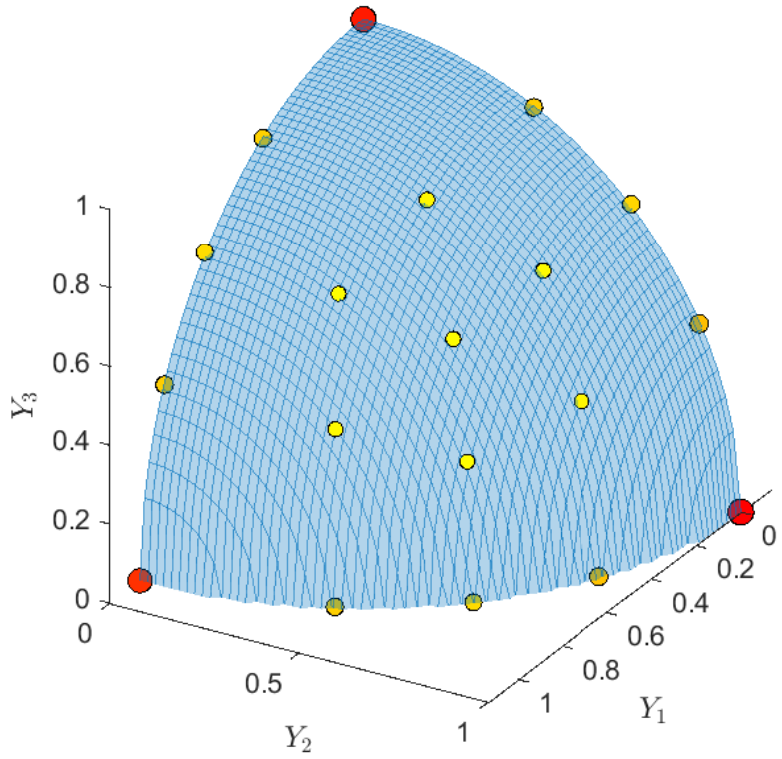

Figure A.3: Experiment 6: 3-dimensional isoquant case: Representation of the directions tested and the values of averaged radial MSE, the size of the markers having a positive affine relation to the values. The color is an other indicator as the more red the higher the averaged radial MSE is and the more yellow for lower values.

We can establish three categories of directions that correspond to certain ranges of MSE values. The first category corresponds to the worst MSE values, which are almost twice the smallest values. These are the directions that have only one non-zero component shown with red markers on the corners of the surface

Table A.5: Experiment 6: Values of radial MSE relative to the true function varying the CNLS-d direction in a 3- dimensional case

CNLS-d Direction $(g^{y_1}, g^{y_2}, g^{y_3})$	Average of radial MSE
$(0, 0, 1)$	9.07
$(0, 0.45, 0.89)$	5.23
$(0, 0.71, 0.71)$	5.04
$(0, 0.89, 0.45)$	5.53
$(0, 1, 0)$	9.62
$(0.33, 0.67, 0.67)$	4.24
$(0.41, 0.41, 0.82)$	4.29
$(0.41, 0.82, 0.41)$	4.35
$(0.45, 0, 0.89)$	5.12
$(0.45, 0.89, 0)$	5.44
$(0.58, 0.58, 0.58)$	4.21
$(0.67, 0.33, 0.67)$	4.15
$(0.67, 0.67, 0.33)$	4.18
$(0.71, 0, 0.71)$	4.89
$(0.71, 0.71, 0)$	4.91
$(0.82, 0.41, 0.41)$	4.23
$(0.89, 0, 0.45)$	5.20
$(0.89, 0.45, 0)$	5.18
$(1, 0, 0)$	8.58

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

shown in Figure A.3. The second category is for the MSE values that are above $5 \cdot 10^{-4}$ but less than $8 \cdot 10^{-4}$. These directions are labeled with the orange markers in Figure A.3 that are on the edges of the surface but not the corners. One of their directional components, $(\boldsymbol{g}^x, \boldsymbol{g}^y)$, is zero but all others are not. The third category of directions, which has the smallest MSEs, correspond to the yellow markers in Figure A.3. These directions have only positive components. Thus, we observe a trend that the directions that have positive components in all variables correspond to the best MSE values. The median value direction, $[0.58, 0.58, 0.58]$, is among the yellow markers. These results support the selection approach proposed in Section 2.6 and confirm the results obtained on the US hospitals data set.

A.3 U.S. Hospital Dataset Application

We describe the functional estimates provided by quadratic regression, CNLS-d using a direction with equal components in all dimensions and CNLS-d using the median direction, and the local linear kernel. Table A.6 provides most productive scale size (MPSS) measurements in cost in $\$M$. Tables A.7 and A.8 provide the marginal cost of Minor Therapeutic procedures and the marginal cost of Major Therapeutic procedures, respectively. The units for Tables A.7 and A.8 are cost in $\$k$ over Minor and Major Therapeutic procedures, respectively.

Our conclusions are the same as stated in the body of the paper, CNLS-d provides the advantage of being more flexible than the parametric estimator (quadratic regression) while having shape constraints that maintain the interpretability of the results.

Table A.6: Most Productive Scale Size (c)

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	529	234	823	105	58	87	109	93	103	228	301	885
25	25	25	50	118	118	122	406	81	369	412	91	112	220	143	216
25	25	25	75	102	110	93	1214	82	895	1220	92	104	209	120	189
25	25	50	25	79	693	560	177	94	166	131	93	140	226	136	162
25	25	50	50	104	139	141	98	60	94	95	79	84	233	217	210
25	25	50	75	105	114	103	165	80	340	179	90	114	219	139	208
25	25	75	25	56	414	335	179	108	554	124	96	384	158	133	27
25	25	75	50	77	245	176	149	93	194	126	91	132	292	117	185
25	25	75	75	94	133	115	93	61	107	105	71	85	226	197	197
25	50	25	25	15	42	63	330	53	78	327	123	127	400	271	1062
25	50	25	50	1074	234	1333	119	55	92	108	91	89	1027	306	215
25	50	25	75	137	131	138	209	78	331	264	98	110	222	153	206
25	50	50	25	248	330	381	80	57	83	93	74	84	373	304	336
25	50	50	50	332	273	1349	70	64	86	127	78	80	903	251	1152
25	50	50	75	133	134	141	177	76	304	182	95	109	233	261	231
25	50	75	25	108	492	718	126	91	144	143	89	129	204	187	79
25	50	75	50	122	694	1068	128	87	137	144	95	112	331	159	239
25	50	75	75	118	154	152	91	59	104	110	77	93	239	232	229
25	75	25	25	11	13	13	915	53	80	1015	125	130	246	98	921
25	75	25	50	11	231	149	192	52	78	197	130	136	537	433	1129
25	75	25	75	1139	223	1542	112	55	75	101	91	115	1015	287	215
25	75	50	25	18	16	5	133	52	79	181	114	118	293	111	887
25	75	50	50	13	311	217	135	51	77	181	111	125	528	466	1091
25	75	50	75	1155	230	1563	109	61	75	99	90	114	1062	272	214
25	75	75	25	64	220	275	81	57	85	94	82	84	300	199	274
25	75	75	50	304	483	484	79	56	93	85	73	83	478	400	437
25	75	75	75	333	265	1532	77	64	96	115	78	79	963	249	153
50	25	25	25	143	189	126	165	115	149	173	157	183	132	139	123
50	25	25	50	119	124	105	126	68	143	110	88	98	287	116	197
50	25	25	75	103	111	90	289	82	424	265	92	104	218	116	185
50	25	50	25	84	740	157	136	72	258	140	91	277	128	209	93
50	25	50	50	106	146	124	96	59	100	101	85	113	245	244	202
50	25	50	75	106	114	100	173	80	292	212	90	113	229	135	211
50	25	75	25	58	431	217	205	97	452	119	95	440	161	128	7
50	25	75	50	79	247	160	140	82	192	114	91	154	150	114	150
50	25	75	75	95	133	111	93	61	106	104	79	106	233	207	197
50	50	25	25	10	142	207	99	51	75	107	111	112	462	363	1031
50	50	25	50	1156	232	1319	109	61	81	114	89	80	1134	367	264
50	50	25	75	138	131	135	208	78	240	267	98	110	233	150	212
50	50	50	25	357	387	450	87	56	90	91	80	90	419	324	194
50	50	50	50	307	272	1329	76	63	84	88	77	78	218	269	652
50	50	50	75	134	135	137	185	76	258	183	95	108	236	170	232
50	50	75	25	110	508	702	125	90	143	124	89	139	209	178	30
50	50	75	50	123	646	1044	128	77	147	119	94	132	333	340	178
50	50	75	75	119	155	149	91	59	103	110	77	103	240	236	236
50	75	25	25	18	15	6	274	53	80	282	124	130	291	117	933
50	75	25	50	14	245	142	191	52	77	188	129	126	566	456	1134
50	75	25	75	1155	224	1523	111	55	75	101	91	115	1050	348	247
50	75	50	25	18	13	10	132	52	79	172	114	118	316	140	894
50	75	50	50	17	325	209	135	51	76	164	111	124	537	502	680
50	75	50	75	1170	230	1544	109	61	83	106	90	114	1106	308	252
50	75	75	25	85	232	264	81	57	84	94	82	84	321	205	245
50	75	75	50	323	493	471	79	56	92	85	81	82	499	406	306
50	75	75	75	335	266	1514	77	64	95	115	78	79	966	252	1192
75	25	25	25	75	101	29	548	309	213	620	177	136	20	27	34
75	25	25	50	100	118	50	129	74	176	142	137	128	100	46	73
75	25	25	75	102	112	81	101	78	133	104	79	99	242	73	160
75	25	50	25	74	142	39	244	95	190	322	92	285	49	57	42
75	25	50	50	95	140	59	112	120	154	189	112	386	123	58	81
75	25	50	75	106	116	83	107	76	131	110	78	99	259	79	179
75	25	75	25	60	534	65	163	75	260	178	84	355	79	115	17
75	25	75	50	80	213	82	139	72	237	179	81	280	129	147	135
75	25	75	75	96	137	96	91	69	111	99	84	114	242	117	188
75	50	25	25	233	593	136	232	128	157	229	254	130	109	542	677
75	50	25	50	185	196	138	171	93	156	145	145	121	154	146	135
75	50	25	75	137	132	115	107	75	118	101	76	106	243	111	182
75	50	50	25	175	670	149	133	85	132	133	98	118	135	278	412
75	50	50	50	169	223	149	120	98	141	121	108	136	179	171	139
75	50	50	75	133	137	118	104	74	117	107	75	95	300	128	211
75	50	75	25	101	607	182	106	71	258	156	80	351	139	161	59
75	50	75	50	117	359	177	102	69	236	150	77	279	171	291	160
75	50	75	75	120	159	131	97	67	108	97	90	111	274	253	210
75	75	25	25	11	57	144	92	52	85	101	92	83	380	220	872
75	75	25	50	12	371	377	88	51	83	105	90	90	642	551	1051
75	75	25	75	740	219	452	101	53	81	86	88	88	253	359	237
75	75	50	25	13	136	202	88	51	84	105	90	92	404	284	866
75	75	50	50	19	439	428	93	50	82	109	89	90	631	630	1048
75	75	50	75	549	225	459	106	59	89	91	88	96	263	349	286
75	75	75	25	296	328	415	79	48	89	85	86	89	363	233	191
75	75	75	50	507	564	593	85	48	88	83	77	97	385	423	191
75	75	75	75	285	261	472	82	56	92	88	76	101	236	264	713

Note: The values displayed are in $M

Table A.7: Marginal Cost of Minor Therapeutic Procedures

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	8.9	6.5	13.2	1.3	1.5	2.5	0.3	0.3	1.5	6.3	10.4	4.7
25	25	25	50	8.9	6.5	13.2	0.1	0.1	0.0	0.3	0.1	0.1	6.1	9.9	3.8
25	25	25	75	8.9	6.5	13.2	0.1	0.0	0.0	0.1	0.0	0.0	5.1	7.8	4.2
25	25	50	25	8.9	6.5	13.2	0.0	0.1	0.4	0.0	0.0	0.6	6.5	10.7	5.9
25	25	50	50	8.9	6.5	13.2	0.1	0.1	0.2	0.0	0.1	0.5	6.4	10.2	5.1
25	25	50	75	8.9	6.5	13.2	0.2	0.0	0.0	0.1	0.0	0.0	5.5	8.0	4.6
25	25	75	25	8.9	6.5	13.2	0.0	0.1	0.1	0.1	0.0	0.6	6.8	10.0	8.2
25	25	75	50	8.9	6.5	13.2	0.0	0.1	0.0	0.0	0.0	0.2	6.8	9.6	7.6
25	25	75	75	8.9	6.5	13.2	0.0	0.0	0.0	0.0	0.0	0.1	5.9	7.8	6.4
25	50	25	25	8.1	6.1	12.4	7.3	8.7	10.3	8.0	8.1	9.6	5.0	10.7	6.2
25	50	25	50	8.1	6.1	12.4	2.8	7.1	8.3	4.9	4.5	8.0	4.8	9.5	4.8
25	50	25	75	8.1	6.1	12.4	1.4	0.2	0.0	0.1	0.0	0.0	4.3	7.0	3.6
25	50	50	25	8.1	6.1	12.4	6.9	5.8	7.7	5.9	5.9	6.0	5.3	10.5	7.8
25	50	50	50	8.1	6.1	12.4	4.1	5.5	7.2	2.3	3.4	6.5	5.2	9.8	6.3
25	50	50	75	8.1	6.1	12.4	0.2	0.0	0.0	0.1	0.0	0.0	4.7	6.9	4.1
25	50	75	25	8.1	6.1	12.4	0.4	1.6	1.2	1.4	0.2	1.7	6.0	9.6	10.6
25	50	75	50	8.1	6.1	12.4	0.5	1.8	0.7	1.4	0.3	0.9	5.9	9.0	9.2
25	50	75	75	8.1	6.1	12.4	0.0	0.0	0.1	0.0	0.0	0.1	5.0	6.7	6.7
25	75	25	25	6.0	5.0	10.4	9.6	13.5	14.0	9.5	11.0	14.2	4.7	8.0	16.0
25	75	25	50	6.0	5.0	10.4	9.6	13.5	14.1	9.6	11.0	14.2	3.8	7.6	14.9
25	75	25	75	6.0	5.0	10.4	5.7	10.1	5.7	4.6	8.6	6.9	3.7	6.3	9.5
25	75	50	25	6.0	5.0	10.4	9.6	13.5	14.1	9.5	10.9	13.8	4.5	7.1	16.5
25	75	50	50	6.0	5.0	10.4	9.6	13.5	14.3	9.6	10.9	13.8	4.0	6.9	15.4
25	75	50	75	6.0	5.0	10.4	5.7	9.6	5.7	4.6	8.1	6.4	3.5	5.8	9.7
25	75	75	25	6.0	5.0	10.4	8.8	12.5	13.1	8.1	10.4	12.2	4.6	7.2	18.4
25	75	75	50	6.0	5.0	10.4	8.8	12.5	13.1	7.8	10.4	12.2	4.3	6.1	17.9
25	75	75	75	6.0	5.0	10.4	4.3	8.9	4.3	2.7	5.8	4.3	3.6	3.6	13.2
50	25	25	25	8.9	6.5	13.2	0.0	0.4	0.1	0.1	0.3	0.2	6.6	10.0	4.9
50	25	25	50	8.9	6.5	13.2	0.1	0.0	0.1	0.1	0.1	0.1	6.4	9.6	4.0
50	25	25	75	8.9	6.5	13.2	0.1	0.0	0.0	0.1	0.0	0.0	5.3	7.9	4.4
50	25	50	25	8.9	6.5	13.2	0.0	0.0	0.0	0.2	0.0	0.1	6.8	10.4	6.1
50	25	50	50	8.9	6.5	13.2	0.0	0.1	0.1	0.0	0.1	0.0	6.7	10.0	5.4
50	25	50	75	8.9	6.5	13.2	0.2	0.0	0.0	0.1	0.0	0.0	5.8	7.9	5.0
50	25	75	25	8.9	6.5	13.2	0.0	0.1	0.0	0.2	0.0	0.1	7.0	9.8	8.6
50	25	75	50	8.9	6.5	13.2	0.0	0.1	0.0	0.1	0.0	0.1	7.1	9.5	7.8
50	25	75	75	8.9	6.5	13.2	0.0	0.0	0.0	0.0	0.0	0.0	6.0	8.2	6.7
50	50	25	25	8.1	6.1	12.4	8.0	8.6	9.7	7.6	6.8	9.9	5.3	10.3	6.6
50	50	25	50	8.1	6.1	12.4	3.9	7.1	7.2	4.9	4.3	7.8	5.1	9.5	5.2
50	50	25	75	8.1	6.1	12.4	1.4	0.4	0.0	0.1	0.0	0.0	4.6	7.2	4.0
50	50	50	25	8.1	6.1	12.4	6.9	5.5	7.4	5.9	6.3	7.8	5.6	10.4	8.0
50	50	50	50	8.1	6.1	12.4	4.3	4.9	7.8	2.1	3.7	7.4	5.5	9.8	6.6
50	50	50	75	8.1	6.1	12.4	0.2	0.4	0.0	0.1	0.0	0.0	4.8	7.2	4.2
50	50	75	25	8.1	6.1	12.4	0.5	1.6	0.8	0.7	0.1	1.0	6.4	9.6	10.2
50	50	75	50	8.1	6.1	12.4	0.5	1.8	0.7	0.6	0.3	0.7	6.3	9.1	9.1
50	50	75	75	8.1	6.1	12.4	0.1	0.0	0.1	0.0	0.0	0.1	5.3	7.1	7.2
50	75	25	25	6.0	5.0	10.4	9.6	13.5	14.0	9.5	11.0	14.2	4.7	7.9	15.9
50	75	25	50	6.0	5.0	10.4	9.6	13.5	14.1	9.6	11.0	14.2	3.9	7.6	13.5
50	75	25	75	6.0	5.0	10.4	5.7	10.1	6.4	4.6	8.7	7.6	3.4	6.4	9.1
50	75	50	25	6.0	5.0	10.4	9.6	13.5	14.0	9.5	10.9	14.1	4.6	7.7	16.7
50	75	50	50	6.0	5.0	10.4	9.6	13.5	14.3	9.6	10.9	13.8	4.1	6.9	15.7
50	75	50	75	6.0	5.0	10.4	5.7	10.1	6.4	4.6	8.7	6.4	3.6	6.0	9.2
50	75	75	25	6.0	5.0	10.4	8.8	12.5	13.1	8.1	10.1	12.2	4.8	7.5	18.4
50	75	75	50	6.0	5.0	10.4	8.8	12.5	13.1	8.2	10.1	12.2	4.3	6.3	17.6
50	75	75	75	6.0	5.0	10.4	4.3	8.9	4.3	2.9	5.8	4.3	3.4	4.4	13.2
75	25	25	25	8.9	6.5	13.2	0.0	0.0	0.3	0.1	0.0	0.3	6.9	9.1	6.9
75	25	25	50	8.9	6.5	13.2	0.2	0.2	0.0	0.5	0.1	0.1	6.6	9.0	6.6
75	25	25	75	8.9	6.5	13.2	0.1	0.1	0.4	0.0	0.0	0.0	6.0	7.9	5.7
75	25	50	25	8.9	6.5	13.2	0.0	0.0	0.3	0.1	0.1	0.1	7.1	9.3	7.8
75	25	50	50	8.9	6.5	13.2	0.2	0.1	0.3	0.3	0.1	0.0	7.0	9.0	7.5
75	25	50	75	8.9	6.5	13.2	0.1	0.1	0.1	0.0	0.0	0.0	6.2	8.0	5.8
75	25	75	25	8.9	6.5	13.2	0.1	0.2	0.3	0.1	0.1	0.2	7.3	8.6	9.5
75	25	75	50	8.9	6.5	13.2	0.1	0.2	0.3	0.2	0.1	0.2	7.1	8.6	8.8
75	25	75	75	8.9	6.5	13.2	0.0	0.1	0.2	0.0	0.0	0.2	6.3	8.1	8.1
75	50	25	25	8.1	6.1	12.4	3.1	2.3	2.9	2.6	1.2	4.0	6.0	9.6	8.4
75	50	25	50	8.1	6.1	12.4	3.0	0.5	3.3	1.7	0.9	1.8	5.9	9.5	7.4
75	50	25	75	8.1	6.1	12.4	0.1	0.1	0.8	0.0	0.2	0.0	5.3	7.9	5.6
75	50	50	25	8.1	6.1	12.4	2.6	2.6	0.4	1.5	2.4	0.5	6.2	9.9	9.2
75	50	50	50	8.1	6.1	12.4	2.1	0.1	0.3	0.8	0.1	0.5	6.2	9.5	8.6
75	50	50	75	8.1	6.1	12.4	0.1	0.1	0.7	0.0	0.2	0.0	5.5	7.7	6.4
75	50	75	25	8.1	6.1	12.4	0.4	0.2	0.5	0.2	0.1	0.8	6.8	8.9	10.8
75	50	75	50	8.1	6.1	12.4	0.4	0.2	0.4	0.2	0.1	0.8	6.7	8.8	10.0
75	50	75	75	8.1	6.1	12.4	0.1	0.1	0.3	0.0	0.0	0.3	5.7	7.8	7.7
75	75	25	25	6.0	5.0	10.4	9.6	13.1	14.4	9.6	11.0	12.6	5.5	8.6	14.8
75	75	25	50	6.0	5.0	10.4	9.6	13.0	14.4	9.6	11.0	12.6	4.8	8.3	14.2
75	75	25	75	6.0	5.0	10.4	4.1	9.0	7.4	3.6	5.6	6.6	3.9	6.9	8.4
75	75	50	25	6.0	5.0	10.4	9.6	13.1	14.4	9.6	11.1	12.5	5.6	8.5	15.5
75	75	50	50	6.0	5.0	10.4	9.6	13.0	14.1	9.6	11.1	12.5	4.9	8.1	15.4
75	75	50	75	6.0	5.0	10.4	4.1	7.6	7.5	3.6	5.6	6.9	3.7	6.6	9.4
75	75	75	25	6.0	5.0	10.4	7.1	8.2	9.5	7.9	6.8	10.7	6.5	8.3	18.1
75	75	75	50	6.0	5.0	10.4	7.1	8.2	9.5	7.9	6.8	10.7	5.6	7.7	17.8
75	75	75	75	6.0	5.0	10.4	4.5	7.7	7.5	3.1	5.3	6.4	4.0	5.3	12.8

Note: The values displayed are in $k

Table A.8: Marginal Cost of Major Therapeutic Procedures

Percentile				Quadratic Regression			CNLS-d (median)			CNLS-d (equal)			LL Kernel		
MinDiag	MinTher	MajDiag	MajTher	2007	2008	2009	2007	2008	2009	2007	2008	2009	2007	2008	2009
25	25	25	25	10.5	11.5	9.8	1.7	2.8	4.4	3.3	1.8	4.9	18.4	14.3	22.9
25	25	25	50	11.7	13.0	10.8	17.3	17.0	20.0	16.8	15.2	18.5	17.5	11.2	18.5
25	25	25	75	15.1	17.2	14.5	19.4	22.3	24.6	19.8	21.8	24.0	15.2	10.2	12.7
25	25	50	25	10.5	11.5	9.8	0.0	0.0	0.2	0.1	0.0	0.1	17.4	14.6	21.7
25	25	50	50	11.7	13.0	10.8	9.6	10.6	13.5	11.2	10.3	13.4	16.8	12.2	18.1
25	25	50	75	15.1	17.2	14.5	19.8	22.2	24.6	19.8	21.8	24.0	15.8	10.5	13.9
25	25	75	25	10.5	11.5	9.8	0.1	0.9	0.1	0.1	0.0	0.2	17.4	14.9	17.2
25	25	75	50	11.7	13.0	10.8	1.3	1.7	1.3	2.9	0.1	5.1	17.3	14.0	17.2
25	25	75	75	15.1	17.2	14.5	16.1	18.0	23.5	16.8	16.5	23.8	16.9	11.8	14.2
25	50	25	25	10.5	11.5	9.8	0.1	0.1	0.2	0.1	0.1	0.5	17.9	12.6	20.0
25	50	25	50	11.7	13.0	10.8	12.9	7.9	8.2	10.0	7.0	6.2	17.1	9.7	16.8
25	50	25	75	15.1	17.2	14.5	19.8	22.3	24.6	19.8	21.8	24.0	15.0	8.7	12.1
25	50	50	25	10.5	11.5	9.8	0.4	0.2	0.4	0.1	0.5	0.3	17.3	13.3	18.9
25	50	50	50	11.7	13.0	10.8	5.2	5.2	1.4	10.5	8.1	5.8	16.6	10.8	16.5
25	50	50	75	15.1	17.2	14.5	19.8	22.2	24.6	19.8	21.8	24.0	15.4	8.5	12.6
25	50	75	25	10.5	11.5	9.8	0.1	0.3	0.1	0.2	0.1	0.2	17.3	14.1	15.7
25	50	75	50	11.7	13.0	10.8	0.1	0.5	0.9	0.2	0.2	4.6	16.9	13.0	16.3
25	50	75	75	15.1	17.2	14.5	16.1	18.0	22.8	16.8	16.5	23.8	15.9	10.2	14.3
25	75	25	25	10.5	11.5	9.8	0.0	0.0	0.1	0.2	0.1	0.1	17.1	9.3	9.9
25	75	25	50	11.7	13.0	10.8	1.6	0.0	0.3	0.7	0.1	0.1	15.4	7.0	9.8
25	75	25	75	15.1	17.2	14.5	18.3	12.4	20.9	16.2	10.9	14.3	15.3	6.2	6.7
25	75	50	25	10.5	11.5	9.8	0.2	0.0	0.2	0.0	0.1	0.2	16.3	9.2	9.4
25	75	50	50	11.7	13.0	10.8	0.2	0.3	0.4	0.8	0.1	0.3	15.6	6.6	8.4
25	75	50	75	15.1	17.2	14.5	18.3	12.8	20.9	16.2	11.3	15.2	15.2	6.2	5.8
25	75	75	25	10.5	11.5	9.8	0.1	0.1	0.1	0.2	0.1	0.1	15.7	11.1	9.8
25	75	75	50	11.7	13.0	10.8	0.1	0.1	0.1	0.6	0.1	0.1	15.6	8.4	9.7
25	75	75	75	15.1	17.2	14.5	15.5	10.4	19.7	17.2	10.8	16.7	14.7	6.9	8.0
50	25	25	25	10.5	11.5	9.8	0.3	0.1	0.0	0.2	0.3	2.7	18.6	14.1	21.4
50	25	25	50	11.7	13.0	10.8	17.8	17.7	17.0	16.3	15.4	19.3	18.0	11.4	17.2
50	25	25	75	15.1	17.2	14.5	19.2	22.3	24.6	19.8	21.8	24.0	15.1	11.1	13.1
50	25	50	25	10.5	11.5	9.8	0.1	0.0	0.1	0.2	0.0	0.1	17.6	14.7	20.4
50	25	50	50	11.7	13.0	10.8	11.3	11.8	15.7	10.5	10.3	14.6	17.2	12.5	17.2
50	25	50	75	15.1	17.2	14.5	19.8	22.1	24.6	19.8	21.8	24.0	15.9	10.7	13.5
50	25	75	25	10.5	11.5	9.8	0.1	1.0	0.3	0.1	0.0	1.3	17.3	15.1	16.8
50	25	75	50	11.7	13.0	10.8	0.9	1.5	2.1	0.5	0.2	1.3	17.2	14.5	16.9
50	25	75	75	15.1	17.2	14.5	16.1	18.0	23.5	16.8	16.5	23.6	16.7	13.0	14.3
50	50	25	25	10.5	11.5	9.8	0.2	0.1	0.2	0.1	0.4	0.5	18.2	12.8	18.6
50	50	25	50	11.7	13.0	10.8	11.1	7.9	10.0	9.4	7.7	5.5	17.6	10.3	15.8
50	50	25	75	15.1	17.2	14.5	19.8	22.2	24.6	19.8	21.8	24.0	14.9	9.4	12.1
50	50	50	25	10.5	11.5	9.8	0.4	0.2	0.5	0.1	0.1	0.4	17.5	13.6	17.6
50	50	50	50	11.7	13.0	10.8	3.7	7.7	1.7	6.9	7.1	3.7	16.9	11.5	15.6
50	50	50	75	15.1	17.2	14.5	19.8	22.0	24.6	19.8	21.8	24.0	15.2	9.5	12.8
50	50	75	25	10.5	11.5	9.8	0.1	0.3	0.2	0.0	0.0	0.4	17.4	14.6	14.8
50	50	75	50	11.7	13.0	10.8	0.1	0.5	0.3	0.1	0.2	1.0	17.0	13.6	15.3
50	50	75	75	15.1	17.2	14.5	17.4	18.0	22.8	16.8	16.5	23.8	16.0	11.5	14.8
50	75	25	25	10.5	11.5	9.8	0.0	0.0	0.1	0.2	0.1	0.0	17.0	9.3	9.5
50	75	25	50	11.7	13.0	10.8	1.6	0.0	0.3	0.7	0.1	0.1	15.7	7.6	6.3
50	75	25	75	15.1	17.2	14.5	18.3	12.4	19.8	16.2	11.0	13.4	14.4	6.6	6.5
50	75	50	25	10.5	11.5	9.8	0.2	0.0	0.1	0.0	0.1	0.1	16.6	10.0	8.9
50	75	50	50	11.7	13.0	10.8	0.2	0.2	0.4	0.8	0.1	0.3	15.8	8.2	8.1
50	75	50	75	15.1	17.2	14.5	18.3	12.4	19.8	16.2	11.0	15.2	14.8	6.7	5.3
50	75	75	25	10.5	11.5	9.8	0.1	0.1	0.1	0.2	0.1	0.1	16.1	11.6	9.6
50	75	75	50	11.7	13.0	10.8	0.1	0.1	0.1	0.3	0.1	0.1	15.5	9.2	9.1
50	75	75	75	15.1	17.2	14.5	15.5	10.4	19.7	16.3	10.8	16.7	14.6	8.3	7.2
75	25	25	25	10.5	11.5	9.8	0.3	0.0	0.3	0.3	0.0	1.5	18.9	14.7	15.0
75	25	25	50	11.7	13.0	10.8	2.4	9.7	4.0	6.7	6.2	4.3	18.0	13.9	13.9
75	25	25	75	15.1	17.2	14.5	19.6	19.5	24.7	19.3	18.3	24.4	15.7	13.2	11.0
75	25	50	25	10.5	11.5	9.8	0.1	0.0	0.2	0.3	0.0	0.2	18.0	15.4	14.8
75	25	50	50	11.7	13.0	10.8	3.9	5.5	0.8	4.5	2.8	3.3	17.6	14.7	13.8
75	25	50	75	15.1	17.2	14.5	19.6	19.5	24.7	19.3	18.9	24.4	16.4	13.4	11.9
75	25	75	25	10.5	11.5	9.8	0.1	0.1	0.3	0.0	0.1	0.4	17.1	16.1	14.0
75	25	75	50	11.7	13.0	10.8	0.1	0.1	0.3	0.2	0.1	0.4	17.1	16.5	14.3
75	25	75	75	15.1	17.2	14.5	19.5	11.6	23.4	19.1	18.5	20.8	17.5	15.4	16.5
75	50	25	25	10.5	11.5	9.8	0.3	0.1	0.5	1.7	0.1	0.7	18.6	13.9	13.2
75	50	25	50	11.7	13.0	10.8	0.9	7.4	1.4	3.1	4.5	3.5	17.8	13.3	12.5
75	50	25	75	15.1	17.2	14.5	19.6	19.5	24.7	19.3	19.2	24.4	15.0	12.4	10.9
75	50	50	25	10.5	11.5	9.8	0.5	0.1	0.2	0.7	0.1	0.1	17.8	15.0	12.8
75	50	50	50	11.7	13.0	10.8	0.7	5.5	0.8	2.5	2.8	3.3	17.3	14.2	12.2
75	50	50	75	15.1	17.2	14.5	19.6	19.5	24.7	19.3	19.8	24.4	15.7	12.3	12.3
75	50	75	25	10.5	11.5	9.8	0.1	0.1	0.2	0.2	0.1	0.2	17.3	16.0	12.1
75	50	75	50	11.7	13.0	10.8	0.1	0.1	0.3	0.2	0.1	0.2	17.0	16.1	12.9
75	50	75	75	15.1	17.2	14.5	19.2	11.6	24.2	19.1	18.5	20.0	16.5	15.0	15.5
75	75	25	25	10.5	11.5	9.8	0.1	0.2	0.1	0.1	0.2	0.3	16.8	11.6	6.9
75	75	25	50	11.7	13.0	10.8	0.1	0.4	0.1	0.1	0.2	0.3	16.2	10.4	6.5
75	75	25	75	15.1	17.2	14.5	18.6	12.6	15.4	15.9	15.0	14.1	14.6	10.1	4.5
75	75	50	25	10.5	11.5	9.8	0.1	0.2	0.1	0.1	0.1	0.1	16.5	12.4	7.2
75	75	50	50	11.7	13.0	10.8	0.1	0.4	0.1	0.1	0.1	0.1	15.8	11.5	7.2
75	75	50	75	15.1	17.2	14.5	18.6	13.4	15.9	15.9	15.0	13.6	14.4	10.6	5.3
75	75	75	25	10.5	11.5	9.8	0.1	0.1	0.1	0.1	0.1	0.2	15.7	14.1	8.5
75	75	75	50	11.7	13.0	10.8	0.1	0.2	0.1	0.1	0.1	0.2	15.3	13.1	7.4
75	75	75	75	15.1	17.2	14.5	13.5	7.2	14.2	12.2	11.7	12.1	14.6	12.7	7.9

Note: The values displayed are in $k

APPENDIX B

APPENDIX OF CHAPTER 3

The appendix is composed of the following parts:

- Formulations of SCBLS when grid is not equal (B.1)
- Additional Results for the US Hospitals Application (B.2)

B.1 Formulations of SCBLS when the knots vectors are not uniform

In the main core of the paper the estimator is introduced for uniform vectors of knots. We write below the formulation for the general case.

B.1.1 SCBLS with coordinate-wise constraints in the univariate case in the non-uniform knots vector case

$$\min_{\gamma_j} \quad \sum_{i=1}^{n} \left(c_i - \sum_{j=1}^{q} \gamma_j B_j^r(Y_i) \right)^2 \tag{B.1}$$

$$\text{subject to} \quad \gamma_j \geq \gamma_{j-1}, \qquad j = 2, \ldots, q \tag{B.1a}$$

$$\frac{\gamma_j}{t_{j+r-1} - t_{j-1}} - 2\gamma_{j-1} \left(\frac{1}{t_{j+r-1} - t_{j-1}} + \frac{1}{t_{j+r} - t_j} \right) + \frac{\gamma_{j-2}}{t_{j+r} - t_j} \geq 0, \quad j = 3, \ldots, q \tag{B.1b}$$

where $\boldsymbol{Y_i},\ c_i,\ i = 1, \ldots, n$ are the observation output levels and cost level respectively, $\gamma_j,\ j = 1, \ldots, q$ are the spline coefficients and $B_j^r,\ j = 1, \ldots, q$ are the basis functions of order $(r+1)$ based on the knots vector $\boldsymbol{t}$.

B.1.2 SCBLS with coordinate-wise constraints in the multivariate case in the non-uniform knots vector case

$$\min_{\gamma_{j_1\dots j_d}} \quad \sum_{i=1}^{n}\left(c_i - \sum_{j_1=1}^{q_1}\dots\sum_{j_d=1}^{q_d}\gamma_{j_1\dots j_d}B^r_{j_1\dots j_d}(\boldsymbol{Y_i})\right)^2 \tag{B.2}$$

$$\text{subject to} \quad \gamma_{j_1\dots j_k\dots j_d} \geq \gamma_{j_1\dots(j_k-1)\dots j_d}, \qquad k=1,\dots,d,\ j_k=2,\dots,q_k$$

$$l=\{1,\dots,d\}\text{-}\{k\},\ j_l=1,\dots,q_l \tag{B.2a}$$

$$\frac{\gamma_{j_1\dots j_k\dots j_d}}{t_{k,(j+r-1)}-t_{k,(j-1)}} - 2\gamma_{j_1\dots(j_k-1)\dots j_d}\left(\frac{1}{t_{k,(j+r-1)}-t_{k,(j-1)}} + \frac{1}{t_{k,(j+r)}-t_{k,j}}\right) + \frac{\gamma_{j_1\dots(j_k-2)\dots j_d}}{t_{k,(j+r)}-t_{k,j}} \geq 0, \qquad k=1,\dots,d,\ j_k=3,\dots,q_k$$

$$l=\{1,\dots,d\}\text{-}\{k\},\ j_l=1,\dots,q_l \tag{B.2b}$$

where $\boldsymbol{Y_i},\ c_i,\ i=1,\dots,n$ are the observation output levels and cost level respectively, $\gamma_{j_1\dots j_d}$ are the spline coefficients and $B^r_{j_1\dots j_d}$ are the products of basis functions as described in Section 3.3.2.2.

B.2 Additional Results for the US Hospitals Application

B.2.1 Estimators performance for a cost function with four outputs.

The dataset is described in the main body of the paper, Section 3.5.1. The estimators applied on the data are the same as in the Base case of the Monte Carlo Simulations (Section 3.4.1.1): 1) SCKLS with coordinate-wise constraints, 2) SCKLS, 3) SCBLS with coordinate-wise constraints, 4) SCBLS with local Afriat constraints and 5) SCBLS with both coordinate-wise and local Afriat constraints. We report the averaged K-fold MSE to the observations, with $k=5$ folds to compare the estimators. The data is equally split into 5 parts. For each part $i=1,\dots,k$, part i is the testing set on which the estimated function is evaluated and the remaining parts are used to determine the estimated function. The averages across folds are reported in Table B.1 alongside the corresponding standard deviations.

In Table B.1, we do not report the values of MSE for SCBLS with only local Afriat constraints as

Table B.1: Averaged K-fold MSE to the observations with its standard deviation (underneath in italics) on two years of the hospital data obtained for different estimators.

Year	SCKLS Coord	SCKLS	SCBLS Coord	SCBLS Afriat+Coord
2007	7.5	7.7	8.0	7.4
	2.6	*2.5*	*2.4*	*2.4*
2008	5.9	6.0	6.1	7.5
	3.3	*3.3*	*3.6*	*5.9*

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

the results obtained are significantly larger than the other values reported. As observed in simulations, Section 3.4, the out-of-sample fitting performance for this estimator is not competitive for cases with more than two regressors. About the results displayed, we notice that the fitting performance is comparable for all estimators in 2007, but that for 2008 the 'SCBLS Afriat+Coord' does slightly worse for than the other estimators for that estimation. Further, the standard deviation of all estimators is high indicating the data set has a high level of noise. In the main body of the paper, Section 3.5.2, we focused on 'SCBLS Afriat+Coord' because it is smooth (a characteristic preferred by some practitioners) and it has sufficient shape constraints to obtain valuable insights about the industry. However, in B.2.3 we report the results of the SCKLS estimator applied to the hospital data. This allows the reader to compare the difference in results due to the piece-wise linear approximation.

B.2.2 Fitting performance with aggregated outputs

We use the same K-fold analysis from B.2.1 to investigate a lower dimensional cost function. The same dataset is used but before estimation, Minor Diagnostic Procedures and Major Diagnostic Procedures are grouped together, and similarly for Minor and Major Therapeutic Procedures. Then a cost function with two regressors (Diagnostic Procedures and Minor Procedures) is estimated. Table B.2 reports the values obtained in this case for the K-fold MSE.

The fitting performances are not very different from one estimator to another in this case. Also contrary to the four-regressor case, 'SCBLS Afriat' also obtain good estimates. This is similar to the results obtained in the simulations section for which all estimators performed well for the two-regressor case and a larger number of observations.

Table B.2: Averaged K-fold MSE to the observations with its standard deviation (underneath in italics) on two years of the hospital data obtained for different estimators.

Year	SCKLS Coord	SCKLS	SCBLS Coord	SCBLS Afriat	SCBLS Afriat+Coord
2007	5.3	5.1	5.2	5.2	5.3
	0.5	*0.6*	*0.5*	*0.5*	*0.4*
2008	6.6	6.7	6.7	6.6	6.7
	2.3	*2.4*	*2.3*	*2.4*	*2.3*

The values displayed are the true values multiplied by 10^3

B.2.3 Results obtained with SCKLS

In this subsection we report the results obtained using 'SCKLS' instead of 'SCBLS Afriat+Coord' as analyzed in the main body of the paper, Section 3.5.2. The Diagnostic procedures being held at their respective median levels, the metrics considered are the marginal rate of substitution between the major and the minor therapeutic procedures (Table B.1), the marginal cost of major therapeutic procedures (Table B.2) and the marginal cost of minor therapeutic procedures (Table B.3). For more details about the metrics and the computation, we refer to Section 3.5.2.

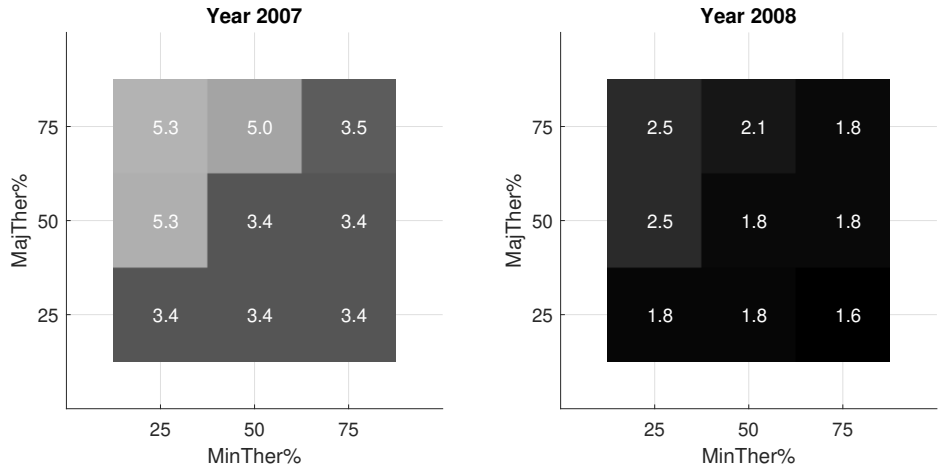

Figure B.1: MRS between Major Therapeutic procedures and Minor Therapeutic procedures obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values.

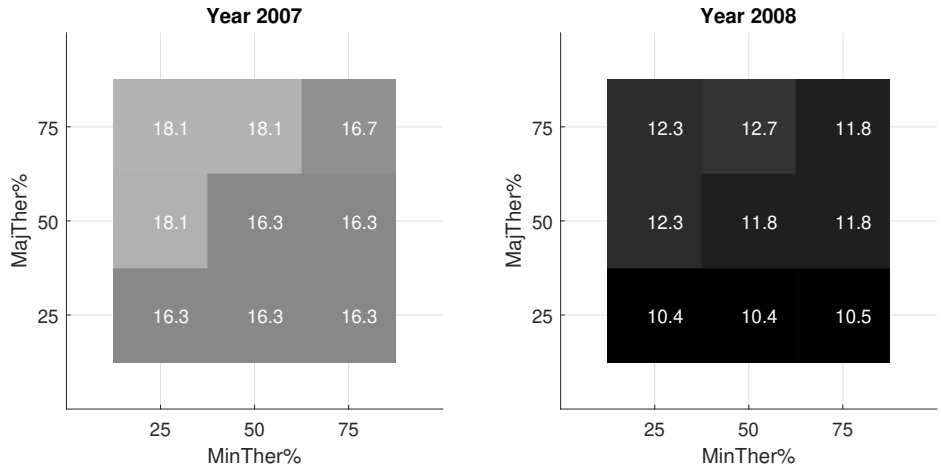

Figure B.2: MC of the Major Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values.

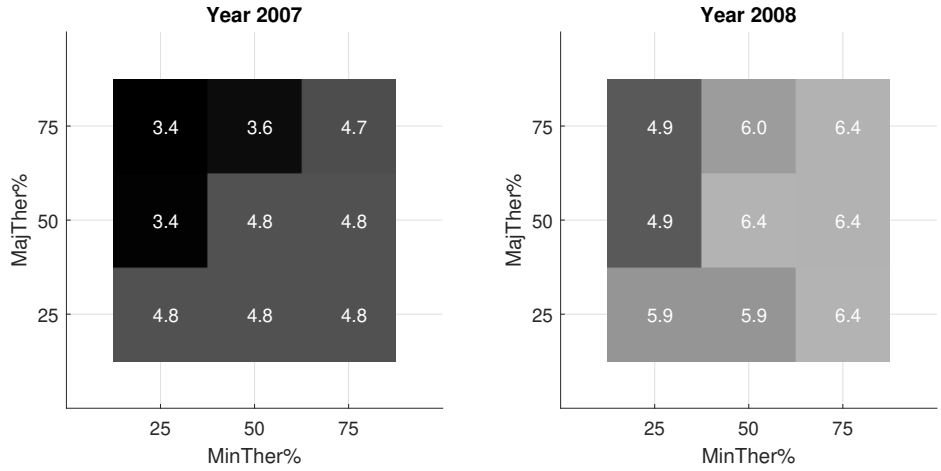

Figure B.3: MC of the Minor Therapeutic procedures (in k$ per procedure) obtained using the estimator 'SCKLS' for years 2007 and 2008, evaluated at different percentile values.

The trends in the results obtained for 'SCKLS' and 'SCBLS Afriat+Coord' are the same. The conclu-

sions from the main body of the paper are confirmed with this estimator. However we can notice that the variations were a lot higher for 'SCBLS Afriat+Coord'. For instance in 2007, the largest MRS is around 8 times the value of the smallest MRS. while for 'SCKLS' it is 1.5 times. Further notice the marginal cost estimated by SCKLS is often constant over multiple percentiles indicating relatively few hyperplanes are used to approximate the cost function. These results indicate that the smoothness (or changes in the marginal cost) observed in the spline estimator are a result of the underlying assumptions of the spline model.