Shape Constrained Nonparametric Estimation of Production Functions

Andrew Johnson
Texas A&M University

Outline

- Introduction/Motivation
- Nonparametric estimation
- Constraint Weighted Bootstrap
  – Constraint Weighted Bootstrap in y-space
- Shape Constrained Kernel-Weighted Least Square
  – Zero-bandwidth estimation
- Smoothness
- Conclusions

Introduction

Shape Constrained Nonparametric Estimation of Production Functions

Introduction

Shape Constrained Nonparametric Estimation of *Production Functions*

Introduction

Shape Constrained Nonparametric Estimation of *Production Functions*

- There is additional structure we can place on production functions coming from production theory
  - Kumbhakar and Lovell (2003)
  - Fare, Grosskopf, and Lovell (1985,1994)
  - Chamber (1988)
  - Varian (1992)

Motivation

To recover the production function

- Regulating cost – cost benchmarks
- Manufacturing – Setting output targets
- Industry characteristics
  - Most productive scale size
  - Returns-to-scale
  - Substitutability of inputs

Nonparametric Methods

Nonparametric methods are asymptotically consistent, so why would you want to impose axioms and risk misspecification?

Nonparametric Methods

Nonparametric methods are asymptotically consistent, so why would you want to impose axioms and risk misspecification?

- Improved finite sample performance
  - Unrestricted nonparametric estimators converge slowly and are typically hard to interpret
  - Parametric assumptions are for statistical/computational convenience, but are not based on production theory and often violate the axioms

Constraint Weighted Bootstrapping

Hall and Huang (2001) and Du et al. (2013)

Characteristics

- Local weighting
- Imposes constraints on a set of grid points
- Estimates a smooth function

Constraint Weighted Bootstrapping

$$\min_{p_j} \sum_{j=1}^{n} (n^{-1} - p_j)^2$$

subject to

$$\hat{g}(\mathbf{x}_i) = \sum_{j=1}^{n} p_j A_j(\mathbf{x}_i) \mathrm{Y}_j \quad 1 \le i \le m$$

$$l(\mathbf{x}_i) \le \hat{g}^{(s)}(\mathbf{x}_i) \le u(\mathbf{x}_i) \quad 1 \le i \le m$$

$$\sum_{j=1}^{n} p_j = 1$$

Constraint Weighted Bootstrapping

Provides a constrained estimate as close as possible to the unconstrained estimator

- If the unconstrained estimator satisfies the constraints, then CWB is not needed
- If the unconstrained estimator does not satisfy the constraints, then this criteria introduces inefficiency in the estimation procedure

Shape Constraints vs. Smoothness

CWB estimates a smooth function

However, CWB only imposes other shape constraints on a set of grid points and results in estimates that do not generally satisfy the constraints globally

CWB can increase the likelihood the shape constraints hold globally by

- Increasing the number of grid points
  - This leads to computational difficulties
- Increasing the bandwidths
  - This leads to over smoothing

CWB in y-space

Provides a constrained estimate as close to the data as possible

- Avoids minimizing the distance to an estimate that is biased in finite samples
- Imposing shape restrictions on a set of grid points

CWB in y-space

$$\min_{p_j} \sum_{j=1}^{n} \Big(y_j - \underbrace{\sum_{i}^{n} p_i A_i(\mathbf{X}_j) y_i}_{=\,\hat{y}_j}\Big)^2$$

subject to

$$\hat{g}(\mathbf{x}_i) = \sum_{j=1}^{n} p_j A_j(\mathbf{x}_i) Y_j \quad 1 \le i \le m \qquad \text{where } A_j(\mathbf{x}_i) \text{ is the local weighting matrix}$$

$$l(\mathbf{x}_i) \le \hat{g}^{(s)}(\mathbf{x}_i) \le u(\mathbf{x}_i) \quad 1 \le i \le m \qquad \text{where } g^{(s)}(\mathbf{x}) = [\partial^{s_1} \ldots \partial^{s_r} g(\mathbf{x})]/[\partial^{s_1} \ldots \partial^{s_r}]$$

$$\sum_{j=1}^{n} p_j = 1$$

Minimize the norm between constrained estimates and observed data

Monte Carlo Simulations

Data generation process

- Cobb-Douglas production function with d-inputs one-output , $y = \prod_{k=1}^{d} x_k^{0.8/d} + \varepsilon$
- Each input x_k is randomly drawn from uniform distribution, $Unif[1,10]$
- Random noise, ε, is sampled from a normal distribution, $N[0, 0.7^2]$
- MSE values are average over 10 trials
- MSE of the CWB estimator in y-space is normalized to 1

#observations		Average of MSE on grid points	
	# grid	100	
	# inputs	2	3
100		8.3	8.7
200		5.7	5.5
300	CWB-p	19.9	7.2
400		13.7	1.6
500		22.8	0.9

#observations		Average of MSE on grid points	
	# grid	300	
	# inputs	2	3
100	CWB-p	12.6	10.2
200		9.4	17.3
300		3.1	4.0
400		24.6	6.7
500		14.2	4.0

#observations		Average of MSE on grid points	
	# grid	500	
	# inputs	2	3
100	CWB-p	12.9	11.0
200		13.1	13.0
300		15.5	12.0
400		6.2	7.9
500		18.3	21.9

Shape Constrained Kernel-Weighted Least Square

Local Linear Kernel Estimator

- Fan (1993) and Fan and Gijbels (1992)

$$\min_{\mathrm{a,b}} \sum_{j=1}^{n} (\mathrm{Y}_j - \mathrm{a} - \left(\mathbf{X}_j - \mathbf{x}\right)' \mathbf{b})^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}}{\mathbf{h}}\right)$$

  - a_i is functional estimate at each grid points
  - $\mathbf{b}_i$ is estimate of slope at each grid points
  - $K\left(\frac{\boldsymbol{X}_j - \boldsymbol{x}_i}{\boldsymbol{h}}\right)$ is the product kernel with bandwidth $\boldsymbol{h}$

Shape Constrained Kernel-Weighted Least Square

$$\min_{\mathrm{a,b}} \sum_{j=1}^{n} (\mathrm{Y}_j - \mathrm{a} - \left(\mathbf{X}_j - \mathbf{x}\right)' \mathbf{b})^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}}{\mathbf{h}}\right)$$

subject to

$$l\left(\mathbf{x}_i\right) \le \hat{g}^{(s)}\left(\mathbf{x}_i, \mathrm{a}, \mathbf{b}\right) \le u\left(\mathbf{x}_i\right) \quad 1 \le i \le m$$

Shape Constrained Kernel-Weighted Least Square

Restricting to monotonic and concave functions

Using a piece-wise linear approximation is

$$\min_{\mathrm{a}_i, \mathrm{b}_i} \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathrm{Y}_j - \mathrm{a}_i - \left(\mathbf{X}_j - \mathbf{x}_i\right)' \mathbf{b}_i)^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}_i}{\mathbf{h}}\right)$$

subject to

$$\mathrm{a}_i - \mathrm{a}_l \ge \mathbf{b}_i'(\mathbf{x}_i - \mathbf{x}_l) \ \ i, l = 1, \ldots, m \ \text{ and } \ i \ne l$$

$$\mathbf{b}_i \ge 0 \quad i = 1, \ldots, m$$

⇒ Concavity (Afriat Inequality)
Monotonicity

Shape Constrained Kernel-Weighted Least Square - Smoothness

- The general Shape Constrained Kernel-Weighted Least Squares estimator results in a smooth estimate of the function
- This approximation is consistent with Afriat (1972) that proposed estimating the minimum function that satisfies monotonicity and convexity (with no additional assumptions of smoothness)

Relationship: CWB (y-space) and Shape Constrained Kernel-Weighted Least Square

Objective Function

- Minimize the sum of squared residuals
- Kernel weight is used to obtain the estimate $\hat{y}_j$

$$\min_{\mathrm{p}_i} \sum_{j=1}^{n} (y_j - \sum_{i}^{n} p_i A_i \left(\mathbf{X}_j\right) y_i)^2 \qquad \text{(CWB-y)}$$

- Minimize the weighted sum of squared residuals
- Kernel weight is used to weight the sum of squared residuals

$$\min_{\mathrm{a}_i, \mathrm{b}_i} \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathrm{Y}_j - \mathrm{a}_i - \left(\mathbf{X}_j - \mathbf{x}_i\right)' \mathbf{b}_i)^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}_i}{\mathbf{h}}\right) \qquad \text{(SCKLS)}$$

Relationships: Shape Constrained Kernel-Weighted Least Square

Assumption 1: The set of evaluation points are equal to observed inputs
i.e. $x_j = X_j \quad j = 1, \dots, n$

Under assumption 1, when the bandwidth approaches to zero, $h \to 0$, SCLWLS estimator is same as Convex Nonparametric Least Squares estimator.

Shape Constrained Kernel-Weighted Least Square

$$\min_{\hat{y}_i, \mathrm{b}_i} \sum_{i=1}^{m} \sum_{j=1}^{n}$$

$$(\mathrm{Y}_j - \hat{y}_i - \left(\mathbf{X}_j - \mathbf{x}_i\right)' \mathbf{b}_i)^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}_i}{\mathbf{h}}\right)$$

subject to

$$\hat{y}_i - \hat{y}_l \geq \mathbf{b}_i'(\mathbf{x}_i - \mathbf{x}_l) \quad i, l = 1, \dots, m \text{ and } i \neq l$$

$$\mathbf{b}_i \geq 0 \quad i = 1, \dots, m$$

Define the grid points as the observations

$h \to 0$

→

Convex Nonparametric Least Squares

$$\min_{\hat{y}_i, \mathrm{b}_i} \sum_{i=1}^{n} \left(\mathrm{Y}_i - \hat{y}_i\right)^2$$

subject to

$$\hat{y}_i - \hat{y}_l \geq \mathbf{b}_i'(\mathbf{x}_i - \mathbf{x}_l) \quad i, l = 1, \dots, n \text{ and } i \neq l$$

$$\mathbf{b}_i \geq 0 \quad i = 1, \dots, n$$

Relationship: Shape Constrained Kernel-Weighted Least Square and CNLS

Seijo and Sen (2011) showed CNLS is an unbiased estimator

Our results show CNLS is the minimum bias estimator in the class of SCKLS estimator

Comments on Smoothness

- An assumption that often has statistical or computational benefits
- Many practitioners like smooth estimators because it is easy to take derivatives
- Non-smooth functions do not have unique derivatives at every point on the function; however, Seijo and Sen (2011) prove for the CNLS estimator, the derivative estimates are consistent
- The non-smooth approximation of Shape Constrained Kernel-Weighted Least Square is computational simpler and has fewer assumption

Monte Carlo Simulations

Data generation process

- d-inputs one-output Cobb-Douglas production function, $y = \prod_{k=1}^{d} x_k^{0.8/d} + \varepsilon$
- Each input x_k is randomly drawn from uniform distribution, $Unif[1,10\,]$
- Random noise, ε, is sampled from a normal distribution, $N[0, 0.7^2]$
- MSE values are average over 10 trials
- MSE of the SCKLS estimator is normalized to 1

# observations		Average of MSE on grid points		
	# grid	100		
	# inputs	2	3	4
100	CNLS	19.0	40.0	44.0
	CWB-p	20.0	13.0	6.6
	OLS	0.4	0.3	0.2
200	CNLS	21.0	32.0	51.0
	CWB-p	17.0	13.0	7.5
	OLS	0.7	0.2	0.2
300	CNLS	32.0	42.0	26.0
	CWB-p	22.0	8.9	7.8
	OLS	0.6	0.3	0.1
400	CNLS	33.0	44.0	52.0
	CWB-p	24.0	11.0	7.9
	OLS	1.1	0.2	0.1
500	CNLS	33.0	72.0	68.0
	CWB-p	22.0	8.3	8.8
	OLS	0.5	0.2	0.1

# observations		Average of MSE on grid points		
	# grid	300		
	# inputs	2	3	4
100	CNLS	19.0	40.0	44.0
	CWB-p	19.0	15.0	6.4
	OLS	0.6	0.5	0.3
200	CNLS	21.0	32.0	51.0
	CWB-p	21.0	19.0	13.0
	OLS	0.6	0.4	0.2
300	CNLS	32.0	42.0	26.0
	CWB-p	17.0	13.0	74.0
	OLS	0.5	0.3	0.2
400	CNLS	33.0	44.0	52.0
	CWB-p	28.0	16.0	10.0
	OLS	0.8	0.4	0.1
500	CNLS	33.0	72.0	68.0
	CWB-p	34.0	13.0	9.8
	OLS	0.8	0.4	0.1

# observations		Average of MSE on grid points		
	# grid	500		
	# inputs	2	3	4
100	CNLS	19.0	40.0	44.0
	CWB-p	18.0	14.0	6.9
	OLS	0.2	0.4	0.4
200	CNLS	21.0	32.0	51.0
	CWB-p	21.0	18.0	12.0
	OLS	0.7	0.4	0.2
300	CNLS	32.0	42.0	26.0
	CWB-p	29.0	16.0	11.0
	OLS	0.5	0.4	0.2
400	CNLS	33.0	44.0	52.0
	CWB-p	20.0	13.0	12.0
	OLS	0.6	0.3	0.1
500	CNLS	33.0	72.0	68.0
	CWB-p	29.0	14.0	13.0
	OLS	1.1	0.3	0.2

# observations		Average of MSE on observation points		
	# grid	100		
	# inputs	2	3	4
100	CNLS	1.3	1.3	1.9
	CWB-p	21.0	16.0	15.0
	OLS	0.5	0.4	0.5
200	CNLS	1.2	1.2	2.2
	CWB-p	20.0	19.0	17.0
	OLS	1.1	0.3	0.4
300	CNLS	1.2	1.1	1.6
	CWB-p	27.0	14.0	19.5
	OLS	0.9	0.6	0.4
400	CNLS	1.1	1.3	1.6
	CWB-p	26.0	14.0	24.0
	OLS	1.5	0.4	0.5
500	CNLS	1.0	1.2	1.5
	CWB-p	27.0	12.0	13.3
	OLS	0.7	0.5	0.3

# observations		Average of MSE on observation points		
	# grid	300		
	# inputs	2	3	4
100	CNLS	1.3	1.3	1.9
	CWB-p	23.0	20.0	16.9
	OLS	0.8	0.7	0.6
200	CNLS	1.2	1.2	2.2
	CWB-p	23.0	22.0	18.3
	OLS	0.6	0.6	0.5
300	CNLS	1.2	1.1	1.6
	CWB-p	17.0	18.0	167.9
	OLS	0.6	0.4	0.5
400	CNLS	1.1	1.3	1.6
	CWB-p	28.0	18.0	62.2
	OLS	0.9	0.6	0.3
500	CNLS	1.0	1.2	1.5
	CWB-p	41.0	16.0	24.9
	OLS	1.1	0.7	0.4

# observations		Average of MSE on observation points		
	# grid	500		
	# inputs	2	3	4
100	CNLS	1.3	1.3	1.9
	CWB-p	19.0	20.0	17.3
	OLS	0.2	0.6	0.7
200	CNLS	1.2	1.2	2.2
	CWB-p	24.0	26.0	17.0
	OLS	0.8	0.6	0.5
300	CNLS	1.2	1.1	1.6
	CWB-p	30.0	20.0	21.4
	OLS	0.6	0.6	0.5
400	CNLS	1.1	1.3	1.6
	CWB-p	21.0	15.0	18.3
	OLS	0.7	0.5	0.3
500	CNLS	1.0	1.2	1.5
	CWB-p	31.0	20.0	33.2
	OLS	1.3	0.6	0.5

Conclusions

- Imposing constraints ex ante on functional estimation is beneficial to improve finite sample estimation
- Local weighting methods use local information to avoid overfitting to the observed sample and are better able to estimate the true function
- Where the kernel weighting enters effects the performance of the estimator

Thank you for your attention