

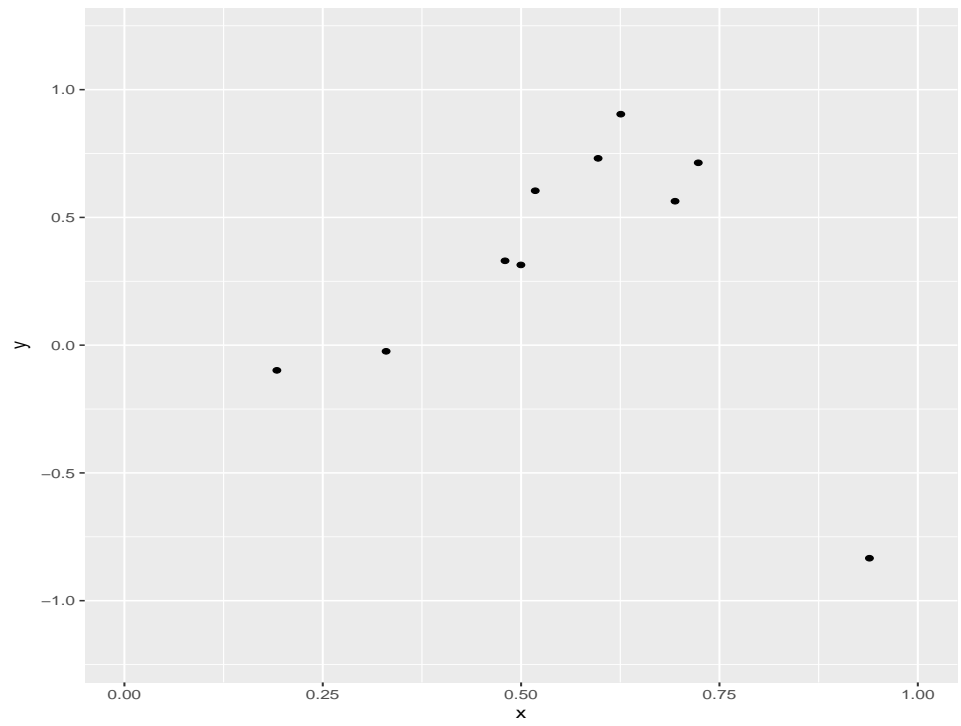
Sieve Estimators

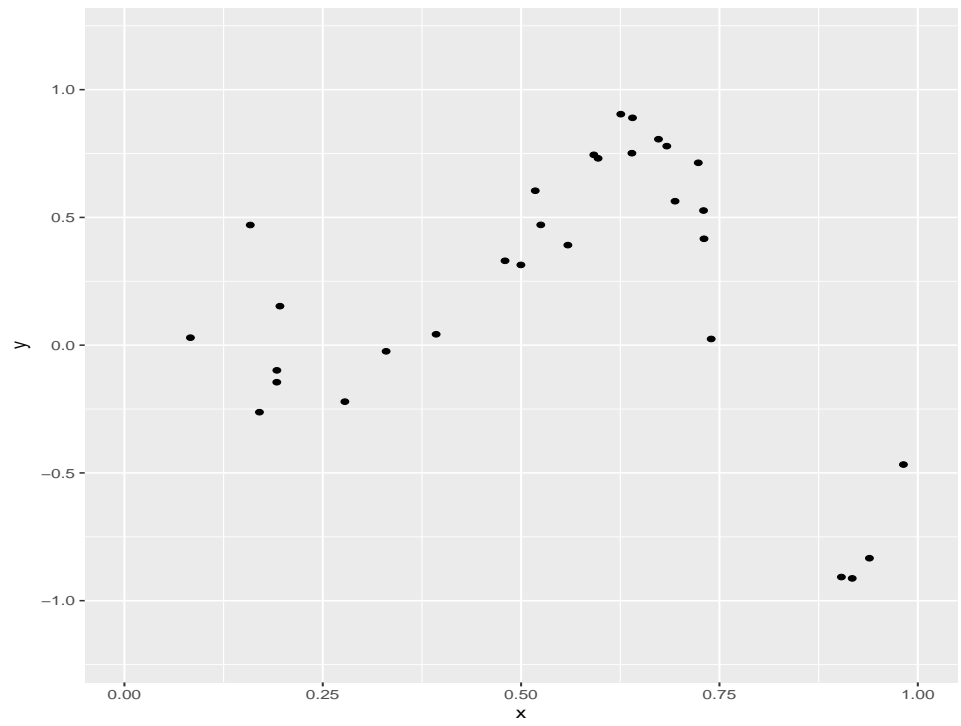
N. Joesphs, M. Wiens, B. Draves

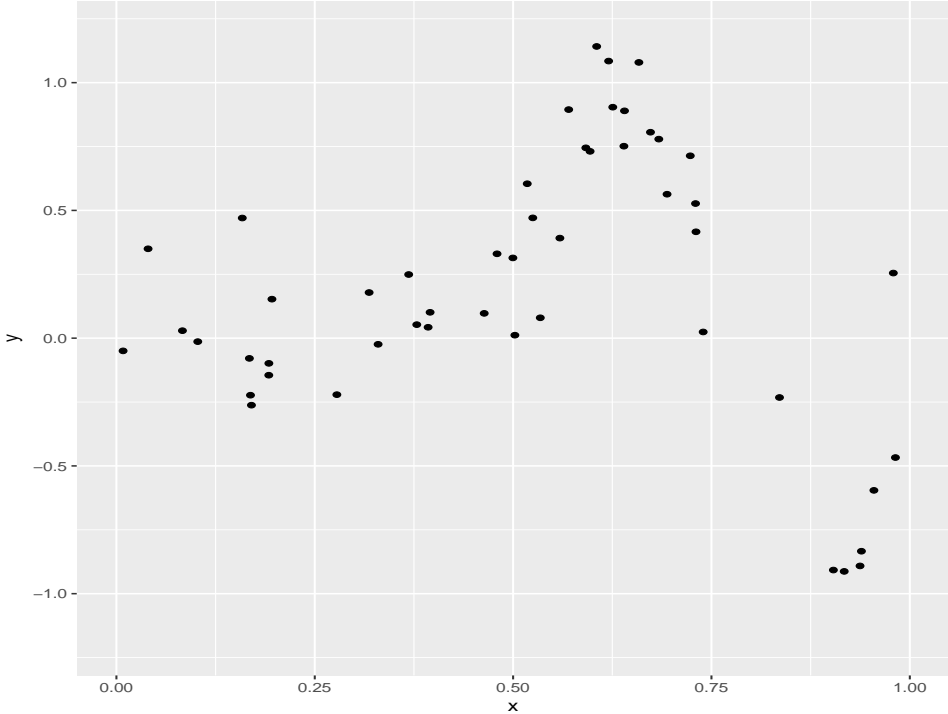
December 7, 2017

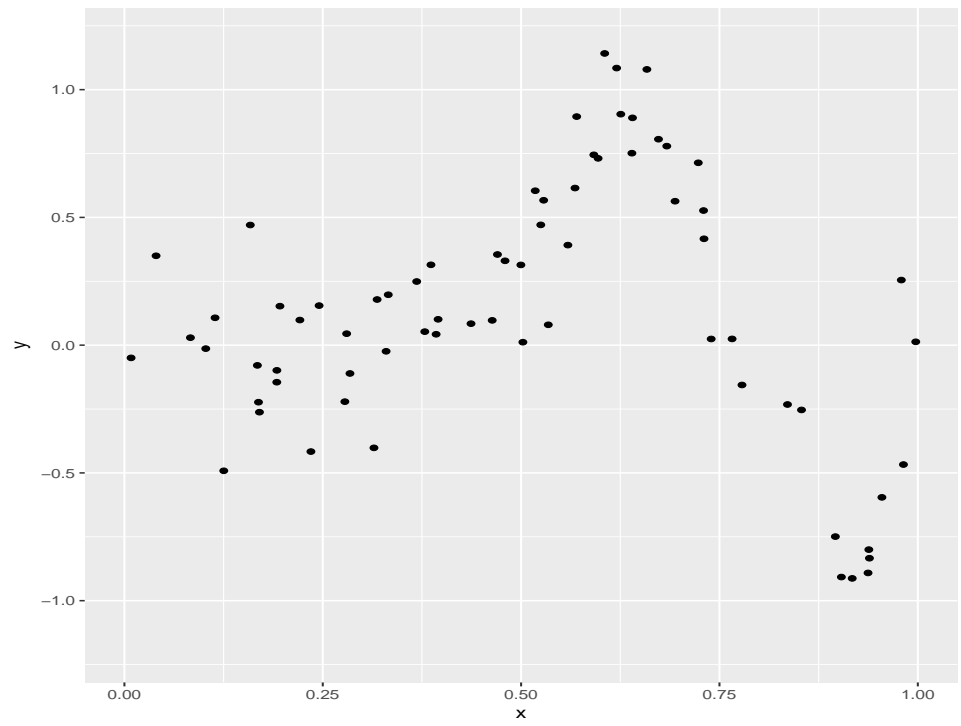
Outline

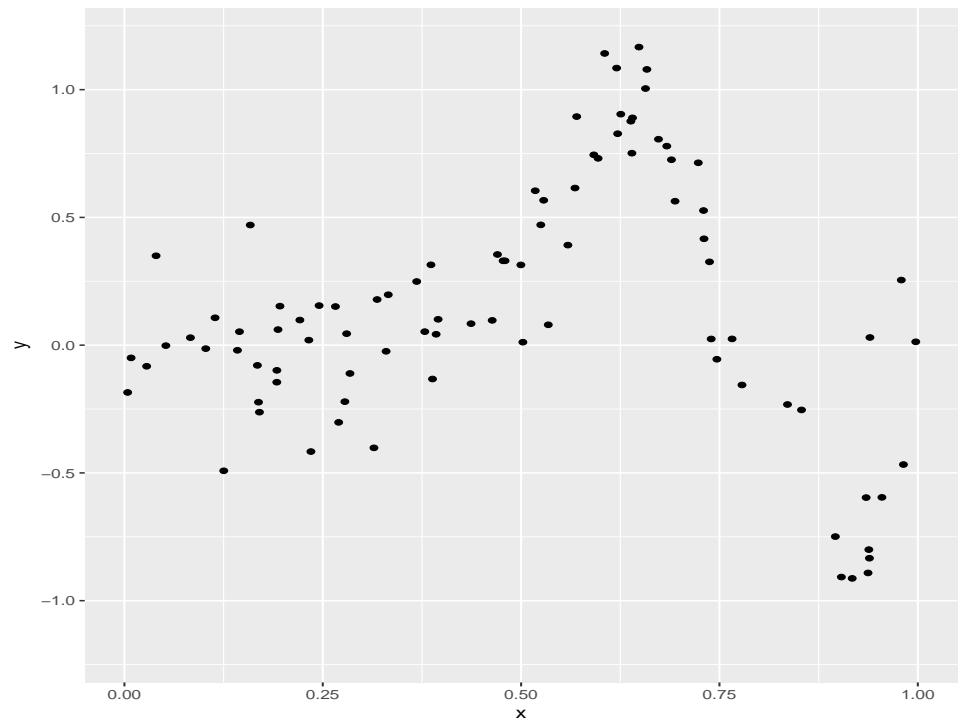
- 1 Define *Sieves Estimator*
- 2 Properties of the Estimator
- 3 Derive Smoothing Selection Algorithm
- 4 Simulations and Applications











Sieves

- *Sieve Estimation* constructs a sequence of function spaces $\{\mathcal{F}_n\}_{n=1}^{\infty}$ such that

$$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \dots$$

- Then, depending on the sample size n , the estimation procedure is performed over the optimal space \mathcal{F}_{opt}
- Heuristically, as n increases, we attain a more robust understanding of the data and should allow the modeling procedure to consider more complex functions

Definitions

- ① **Goal:** Estimate unknown regression function for $Y = f(X)$
- ② We consider the *series estimators* of f

$$\mathcal{F} = \left\{ g(x) : g(x) = \sum_{d=0}^D \alpha_d \phi_d(x) \right\}$$

where $D \rightarrow \infty$ and $n \rightarrow \infty$

Definitions

- ① **Goal:** Estimate unknown regression function for $Y = f(X)$
- ② We consider the *series estimators* of f

$$\mathcal{F} = \left\{ g(x) : g(x) = \sum_{d=0}^D \alpha_d \phi_d(x) \right\}$$

where $D \rightarrow \infty$ and $n \rightarrow \infty$

- $\phi_d(x)$ is the d th basis function
 - α_d is the corresponding coefficient
- ③ The complexity of \mathcal{F} increases as n increases

Basis Functions

- ① Choosing basis function is analogous to choosing kernel
- ② Basis functions should represent belief about the underlying function
- ③ Some popular choices include wavelets and splines and others given below

Basis Functions	Functional Form
Polynomials	$\sum_{i=0}^d c_d x^d$
Fourier	$a \cos(\pi dx) + b \sin(\pi dx)$
Gaussian	$\varphi^{(d)}(x)$

Table: Series estimators with basis function ϕ for $\hat{g}(x)$

OLS Formulation

- 1 We note the similarity between the series estimator and a general linear estimator

OLS Formulation

- ① We note the similarity between the series estimator and a general linear estimator
- ② Using the following matrices

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, P_D = \begin{bmatrix} \phi_0(x_1) & \dots & \phi_D(x_1) \\ \vdots & \ddots & \vdots \\ \phi_0(x_n) & \dots & \phi_D(x_n) \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_D \end{bmatrix}, e = \begin{bmatrix} e(x_1) \\ e(x_2) \\ \vdots \\ e(x_n) \end{bmatrix}$$

we can write our model as

$$Y = P_D \alpha + e$$

- ③ Bias and variance depend on choosing D correctly
- ④ Note that D is our smoothing parameter - analogous to the bandwidth in kernel density estimation

MISE

- ① We choose our optimal dimension D_{opt} by minimizing MISE
- ② We can decompose the error of the estimate as $e(x_i) = \epsilon(x_i) + r(x_i)$
 - $\epsilon(x_i)$ random error around true regression function
 - $r(x_i)$ error due to misspecification of D

MISE

- ① We choose our optimal dimension D_{opt} by minimizing MISE
- ② We can decompose the error of the estimate as $e(x_i) = \epsilon(x_i) + r(x_i)$
 - $\epsilon(x_i)$ random error around true regression function
 - $r(x_i)$ error due to misspecification of D

- ③ MISE can be written as

$$MISE(D) = E[r^2(x)] + tr \left[E[\phi(x)\phi(x)^T] E \left((\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T \right) \right]$$

- ④ Even under the assumption of homoscedasticity

$$MISE^*(D) \simeq E[r^2(x)] + \frac{\sigma^2 D}{n}$$

which is still a function of the unknown $r(x_i)$

Connecting MISE and PSE

- 1 While MISE is infeasible, MISE is connected to PSE
- 2 Let x^* be a new data point. The sieve estimate of x^* is given by

$$y^* = \hat{g}(x^*)$$

with error e^*

- 3 Then we can show the relation

$$PSE(\hat{g}(x^*)) = \text{Var}(e^*) + MISE(\hat{g}(x))$$

- 4 Therefore, minimizing MISE is equivalent to minimizing PSE

Connecting PSE to CV

- 1 Let $\tilde{e} = Y^* - \hat{y}^*$ then $PSE(\hat{g}(x^*)) = E[\tilde{e}^2]$
- 2 We interpret this as the expectation of a single leave-one-out squared prediction error

Connecting PSE to CV

- 1 Let $\tilde{e} = Y^* - \hat{y}^*$ then $PSE(\hat{g}(x^*)) = E[\tilde{e}^2]$
- 2 We interpret this as the expectation of a single leave-one-out squared prediction error
- 3 This gives rise to the leave-one-out cross validation criterion

$$CV(\hat{g}) = \frac{1}{n} \sum_{i=1}^n \tilde{e}_i^2$$

where $\tilde{e}_i = y_i - \hat{y}_{(i)}$

- 4 This establishes a relation between MISE, PSE, and CV

Connecting PSE to CV

- 1 Let $\tilde{e} = Y^* - \hat{y}^*$ then $PSE(\hat{g}(x^*)) = E[\tilde{e}^2]$
- 2 We interpret this as the expectation of a single leave-one-out squared prediction error
- 3 This gives rise to the leave-one-out cross validation criterion

$$CV(\hat{g}) = \frac{1}{n} \sum_{i=1}^n \tilde{e}_i^2$$

where $\tilde{e}_i = y_i - \hat{y}_{(i)}$

- 4 This establishes a relation between MISE, PSE, and CV
- 5 Hansen (2012) show that choosing D based on CV is asymptotically equivalent to choosing D based on MISE

Simulation Design

- 1 Recall the model from HW5 given by

$$Y = \sin^3(2\pi X^3) + \epsilon$$

where

$$X \sim N(0, 1)$$

$$\epsilon \sim N(0, .2)$$

- 2 Our previous solution: LLR and Kernel Regression
- 3 Applying polynomial sieves
 - $N = 150$ data points
 - From $n = 10$ to N in increments of 10, we fit polynomial series estimator with degrees of complexity varying from 1 to 8
 - Repeated this on 200 separate datasets and plotted the PSE against n
- 4 Simulation Results

Geyser Data Application

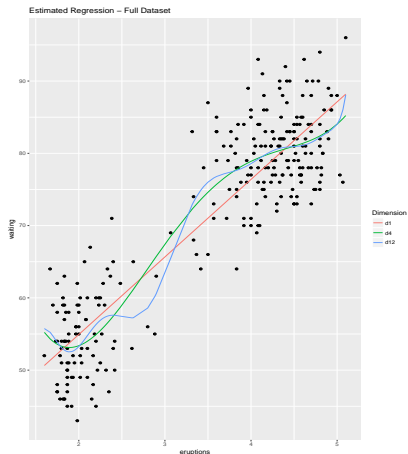
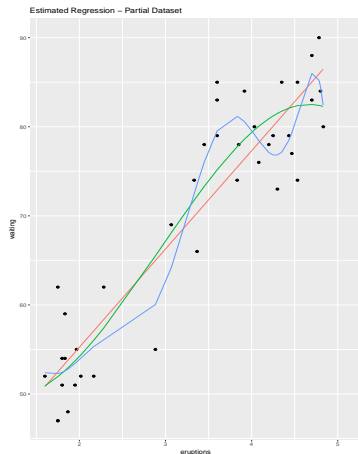


Figure: Sieves Analysis of Geyser dataset. We see that increasing our sample size leads to differing optimal models

Geyser Data Application

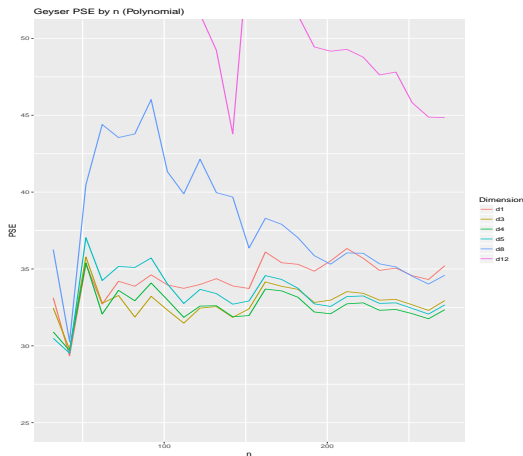


Figure: Geyser data PSE by n

Sieve Estimators and Monetary Policy

- 1 Sa and Portugal (2015) attempt to estimate a multi-dimensional loss function with a sieve estimator
- 2 The authors use a polynomial basis to estimate derivatives
- 3 Based on tests for third derivative they conclude there is asymmetry in the loss function
- 4 This corresponds to additional loss for low inflation (deflation) with regard to deciding policy

Conclusion

- 1 Introduced sieves as a nonparametric dimension adaptive estimator

Conclusion

- 1 Introduced sieves as a nonparametric dimension adaptive estimator
- 2 Analyzed and developed for the case of polynomial series estimators in regression

Conclusion

- 1 Introduced sieves as a nonparametric dimension adaptive estimator
- 2 Analyzed and developed for the case of polynomial series estimators in regression
- 3 Showed a data driven process for selecting D

Conclusion

- 1 Introduced sieves as a nonparametric dimension adaptive estimator
- 2 Analyzed and developed for the case of polynomial series estimators in regression
- 3 Showed a data driven process for selecting D
- 4 Illustrated the estimators in simulation and real data applications