NONPARAMETRIC REGRESSION WIT

MEASUREMENT ERROR: SOME RECENT PR

David Ruppert

Cornell University

www.orie.cornell.edu/~davidr

(These transparencies, preprints, and references a

link to “Recent Talks” and “Recent Paper

Work done jointly with

Scott Berry, Berry Consultants

R. J. Carroll, Texas A & M

Jeff Maca, Novartis

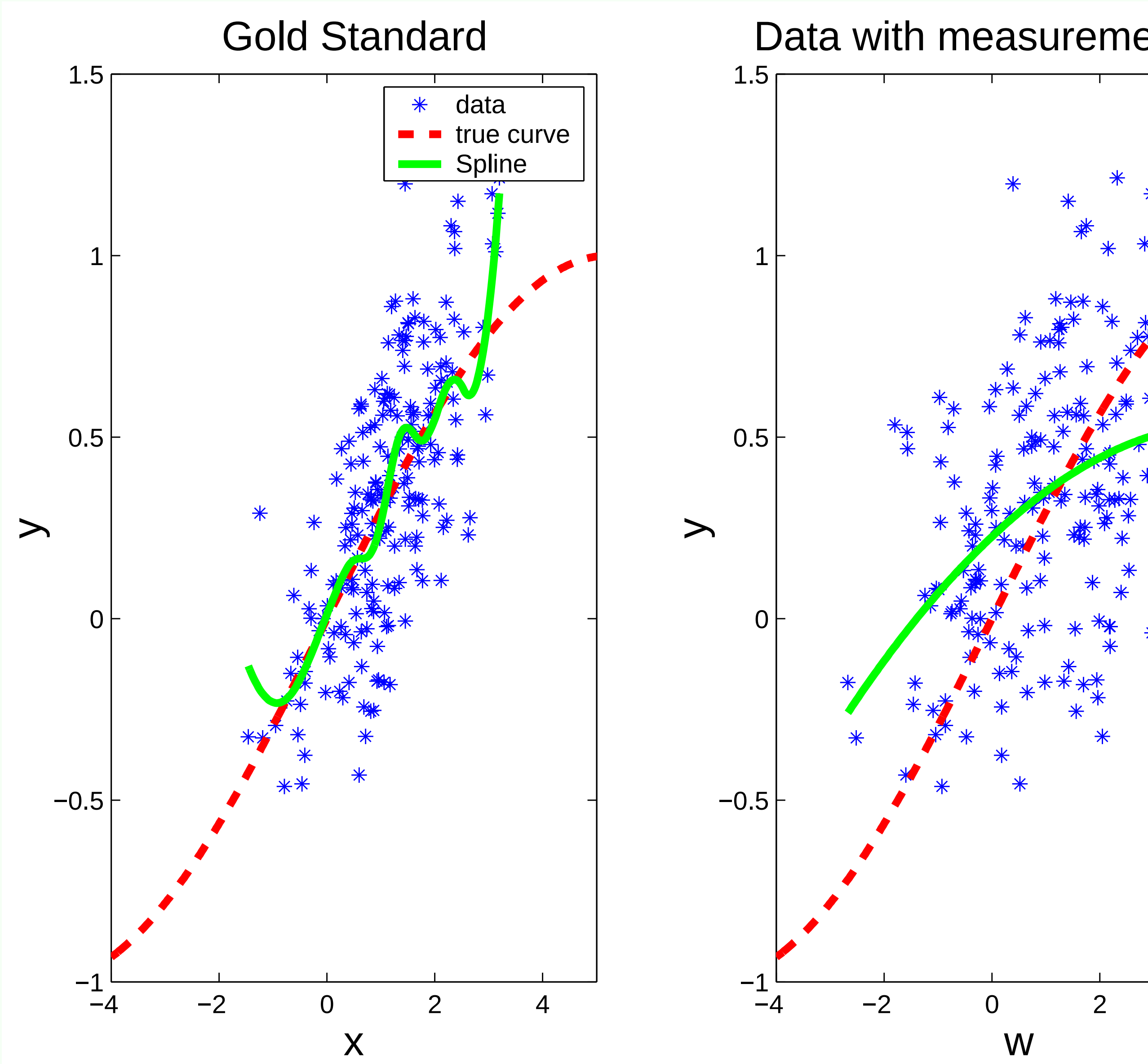
John Staudenmayer, U Mass

OUTLINE

- The problem — nonparametric regression with error
- Review of the currently available estimators
- New Bayesian spline approach (Berry, Carroll, 2002, JASA)
- Simulation results

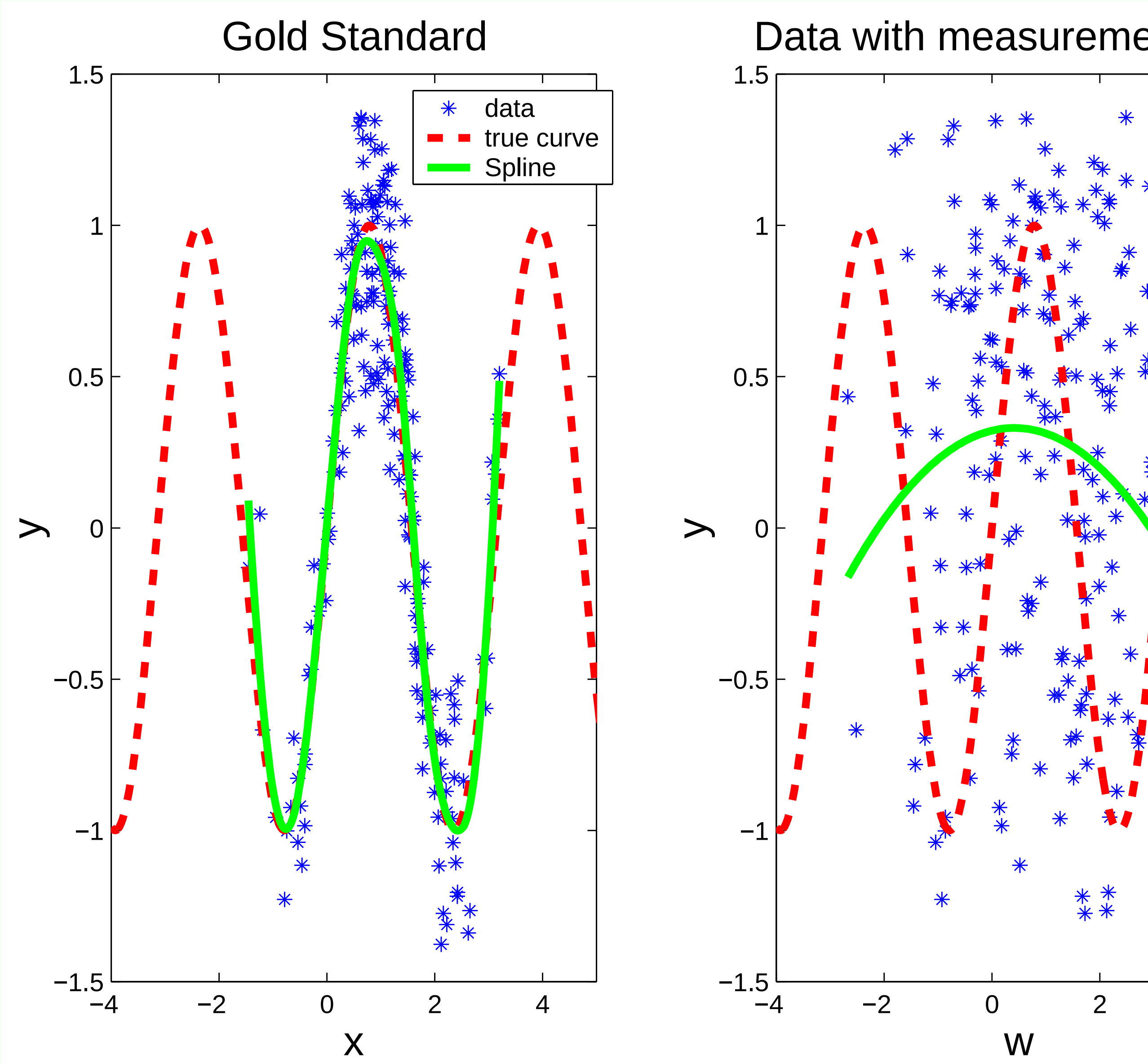
THE PROBLEM OF MEASUREMENT ERR

ILLUSTRATION



THE PROBLEM OF MEASUREMENT ERR

ILLUSTRATION



THE PROBLEM OF MEASUREMENT ER

- The regression model is

$$Y = m(X) + \epsilon$$

  where m is only known to be smooth

- Observe

$$Y \text{ and } W = X + U, \text{ where}$$

  - $E(U|X) = 0$
  - var($U|X$) = σ_u^2
  - $U|X$ normally distributed

- Measurement error variance σ_u^2 is estimated from cate data. (Observe W_{ij}, $j = 1, \ldots, n_i$.)

THE PROBLEM OF MEASUREMENT ERRO

- Measurement error occurs in a wide variety of pro
  - Measuring nutrient intake
  - Measuring airborne lead exposure
  - Measuring blood pressure
  - C_{14} dating
- The effects of measurement error are:
  - biased estimates of the regression curve
  - increase in the perceived variability about the re

THE PROBLEM OF MEASUREMENT ERRO

- Other than the work of Fan and Truong (1993, A
  had been little done on nonparametric regression w
  ment error until
  - Carroll, Maca, and Ruppert (1999, *Biometrika*) (
  - Berry, Carroll, Ruppert (2002, *JASA*) (**BCR**)

REVIEW OF CURRENT ESTIMATORS

- Globally consistent nonparametric regression by kernels (Fan and Truong, 1993, *Annals*)
  - does not work so well
    - ∗ Fan & Truong show very poor asymptotic rat gence
    - ∗ we have simulations showing poor finite-samp
  - no methodology for bandwidth selection or infer

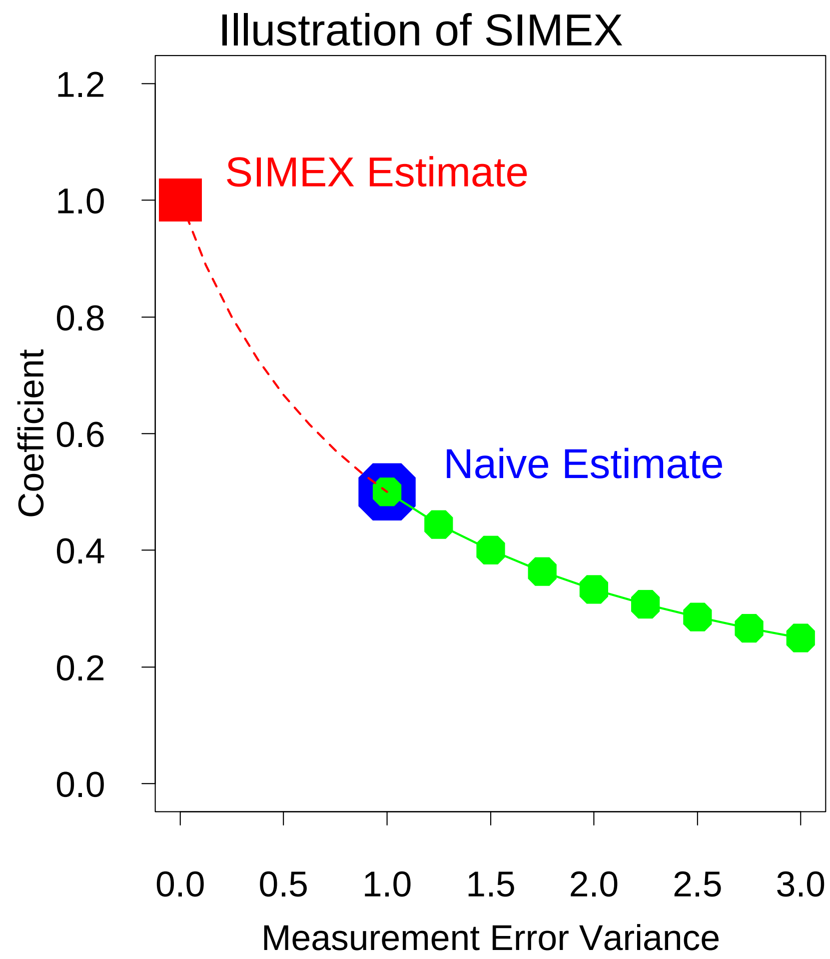
REVIEW OF CURRENT ESTIMATORS

- Standard measurement error method: SIMEX
  - functional — no assumptions on $[X]$
  - very general — can be applied to nearly any error problem, parametric or nonparametric
- Structural Spline
  - Regression splines for basic regression model
  - Mixtures of normals for covariate density model
  - Emphasis is on flexible parametric modeling, no ric modeling. (Little or no difference in practice.)

SIMEX

- The SIMEX method is due to Cook & Stefanski (1
  - The theory is in Carroll, et al. (1996, *JASA*)
  - Also see Carroll, Ruppert, and Stefanski (1995, *Error in Nonlinear Models*)
- SIMEX has been previously applied to parametric
  - makes no assumptions about the true X's. (*Fur*
  - results in estimators which are *approximately* co consistent at least to order $O(\sigma_u^6)$.
- Here is the method, defined via a graph.

SIMEX, ILLUSTRATED



SIMEX

- **CMR** applied the SIMEX to nonparametric regres
- **CMR** have asymptotic theory in the local polynom (LPR) context.
  - The estimators have the usual rates of converge
  - They are approximately consistent, to order $O($
- An asymptotic theory with rates seems very difficu
  - but, simulations in **CMR** indicate that SIMEX/sp little *better* than SIMEX/kernel
  - problem seems due to undersmoothing

SIMEX

- Staudenmayer (2000, Cornell PhD thesis) looked
  selection for SIMEX/LPR.
  - With better bandwidth selection, SIMEX/LPR i
    with other methods.

- The regression of Y on the observed W is

$$E(Y|W) = E\{m(X)|W\} = \int m(x) f(x|W$$

- If we had a model $m(X;\boldsymbol{\beta})$ for $m(X)$and if we knew
we could estimate $m(X;\boldsymbol{\beta})$ by minimizing over the

$$\sum_{i=1}^{n} \left\{ Y_i - \int m(x;\boldsymbol{\beta}) f(x|W_i)\, dx \right\}^2.$$

- We need two things to make this work:
  - convenient flexible form for $m(x;\boldsymbol{\beta})$
  - convenient flexible distribution for X.

REGRESSION SPLINES

- Model

$$E(Y|X) = m(X; \boldsymbol{\beta}) := \sum_{j=0}^{J} \beta_j X^j + \sum_{j=1}^{K} \beta_{j+J}(X$$

- The key remaining issue: the joint distribution of X

  – **CMR** used a mixtures of normals for $[X]$ and G
  to estimate the parameters.

    * This is an extension to measurement error
    Roeder & Wasserman (*JASA*, 1997).

What's New?

Answer: Fully Bayesian MCMC method in **BCR**

- Uses splines
  - smoothing or penalized
  - P-splines in this talk
- Structural
  - X_i are iid normal
  - but seems robust to violations of normality

FULLY BAYESIAN MODEL

- Smoothing parameter is automatic
- Inference adjusts for the data-based smoothing p
  for measurement error
- Allow global confidence bands

FULLY BAYESIAN MODEL — PARAMET

- **Regression Model**

  $Y_i = m(x_i; \boldsymbol{\beta}) + \epsilon_i$

  – $m(x_i; \boldsymbol{\beta})$ is a P-spline

  – ϵ_i iid $N(0, \sigma_\epsilon^2)$

- **Measurement Error Model**

  $W_{ij} = X_i + U_{ij}$ where U_{ij} iid $N(0, \sigma_u^2)$

- **Structural Model**

  X_i iid $N(\mu_x, \sigma_x^2)$

- **Parameters:** $\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2, \mu_x, \sigma_x^2$

FULLY BAYESIAN MODEL — PARAMET

- **Priors**
  - $\boldsymbol{\beta}$ is $N(0, (\gamma \boldsymbol{K})^{-1})$ where $\boldsymbol{K}$ is known. [$\alpha := \gamma\sigma_e^2$ i
    ing parameter.]
  - γ is Gamma(A_γ, B_γ)
  - σ_e^2 is Inv-Gamma(A_e, B_e)
  - σ_u^2 is Inv-Gamma(A_u, B_u)
  - μ_x is $N(d_x, t_x^2)$
  - σ_x^2 is Inv-Gamma(A_x, B_x)
- **Hyperparameters**: $A_e, B_e, A_u, B_u, A_x, B_x, d_x, t_x^2, A_\gamma$
  - all fixed at values making the priors noninforma
    - E.g., $t_x^2 = 10^6$.

GIBBS SAMPLING

- Iterate through $\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2, \sigma_x^2, \mu_x, \gamma, X_1, \ldots, X_n$.
- All steps except one are easy, either gamma, inv
  or normal
  - E.g.,

$$[\boldsymbol{\beta}|\text{other parameters}, \boldsymbol{Y}, \boldsymbol{W}] \sim \text{Normal}$$
$$\text{Mean} = (\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} + \gamma\boldsymbol{K})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y}$$
$$\text{Cov} = \sigma_e^2(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} + \gamma\boldsymbol{K})^{-1}.$$

    * Here $\boldsymbol{X}$ is one of the "other parameters"
    * Essentially we're fitting a spline to the impute
      observed Y's

* Estimate of $\boldsymbol{\beta}$, call it $\widehat{\boldsymbol{\beta}}$, is

$$(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X} + \gamma \boldsymbol{K})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y}$$

averaged over γ and $\boldsymbol{X}$.

GIBBS SAMPLING

- The exception to the sampling being quick and c
  Metropolis-Hastings step is needed for $X_1, \ldots, X_n$

$$[X_i | \mu_x, \sigma_x^2, \boldsymbol{\beta}, \sigma_e^2, \sigma_u^2, \boldsymbol{Y}, \mathbf{W}]$$
$$\propto \exp(-\frac{1}{2\sigma_u^2} \sum_{j=1}^{m_i} (W_{ij} - X_i)^2$$
$$-\frac{1}{2\sigma_\epsilon^2}\{Y_i - m(X_i; \boldsymbol{\beta})\}^2 - \frac{1}{2\sigma_x^2}(X_i - \mu_x)^2$$

- Let $\tilde{\boldsymbol{X}}$ be the spline basis function evaluated on a
  some interval, $[a, b]$.
- $\tilde{\boldsymbol{X}}\boldsymbol{\beta}$ is the curve on $[a, b]$..
- $\tilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}}$ is the estimated curve.
- Let K_α be the $(1 - \alpha)$ MCMC sample quantile of

$$\max_{\text{grid}} \left\{ \frac{\tilde{\boldsymbol{X}}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})}{\text{SD}(\tilde{\boldsymbol{X}}\boldsymbol{\beta})} \right\}.$$

- Then,

$$\tilde{\boldsymbol{X}}\widehat{\boldsymbol{\beta}} \pm K_{.95}\,\text{SD}(\tilde{\boldsymbol{X}}\boldsymbol{\beta})$$

  is a 100$(1-\alpha)$% simultaneous confidence band fo
  $[a, b]$.

- Let $\tilde{\boldsymbol{X}}'$ be derivatives of the spline basis function e
  fine grid over $[a,b]$.
- $\tilde{\boldsymbol{X}}'\boldsymbol{\beta}$ is the curve's derivative on $[a,b]$.
- $\tilde{\boldsymbol{X}}'\widehat{\boldsymbol{\beta}}$ is the estimated derivative.
- Let K'_α be the $(1-\alpha)$ MCMC sample quantile of

$$\max_{\text{grid}} \left\{ \frac{\tilde{\boldsymbol{X}}'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})}{\mathrm{SD}(\tilde{\boldsymbol{X}}'\boldsymbol{\beta})} \right\}.$$

- Then,

$$\tilde{\boldsymbol{X}}'\widehat{\boldsymbol{\beta}} \pm K'_{.95}\,\mathrm{SD}(\tilde{\boldsymbol{X}}'\boldsymbol{\beta})$$

  is a 100$(1-\alpha)$% simultaneous confidence band for
  on $[a,b]$.

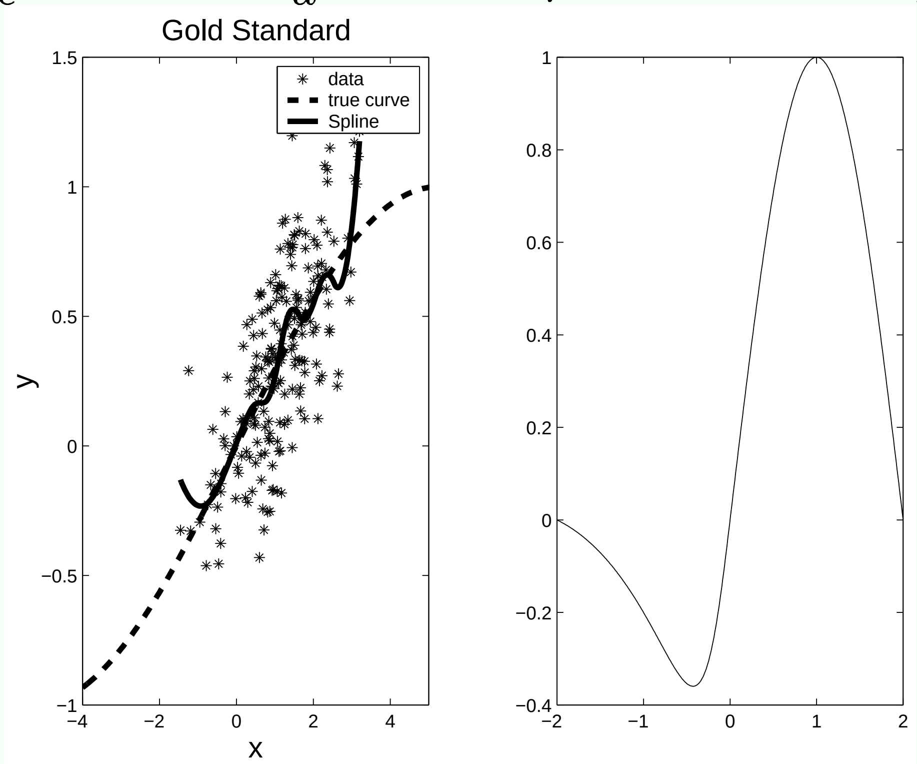
SIMULATIONS

The six cases were considered. $n_i \equiv 2$ in each case

Case 1 The regression function is

$$m(x) = \frac{\sin(\pi x/2)}{1 + 2x^2\{\text{sign}(x) + 1\}}.$$

with $n = 100$, $\sigma_\epsilon^2 = 0.3^2$, $\sigma_u^2 = 0.8^2$, $\mu_x = 0$ and $\sigma_x^2 =$



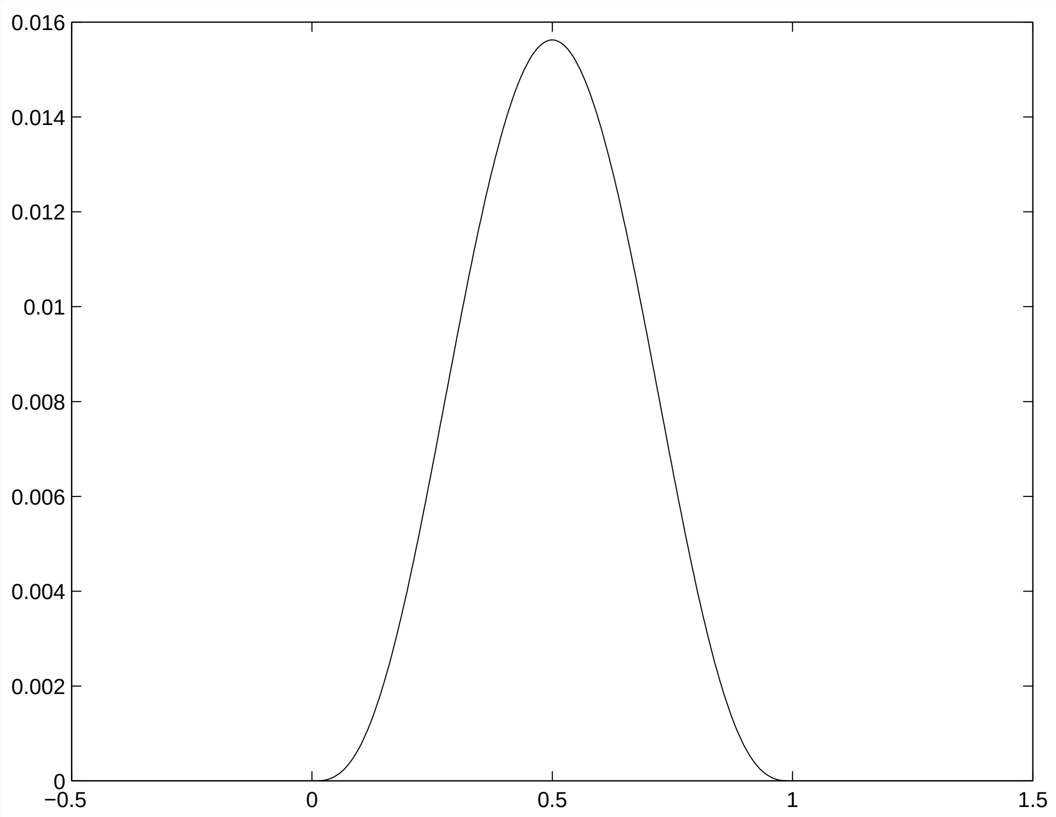
Case 2 Same as Case 1 except $n = 200$.

Case 3 A modification of Case 1 above except that n

Case 4 Case 1 of **CMR** so that

$$m(x) = 1000x_+^3(1-x)_+^3,$$

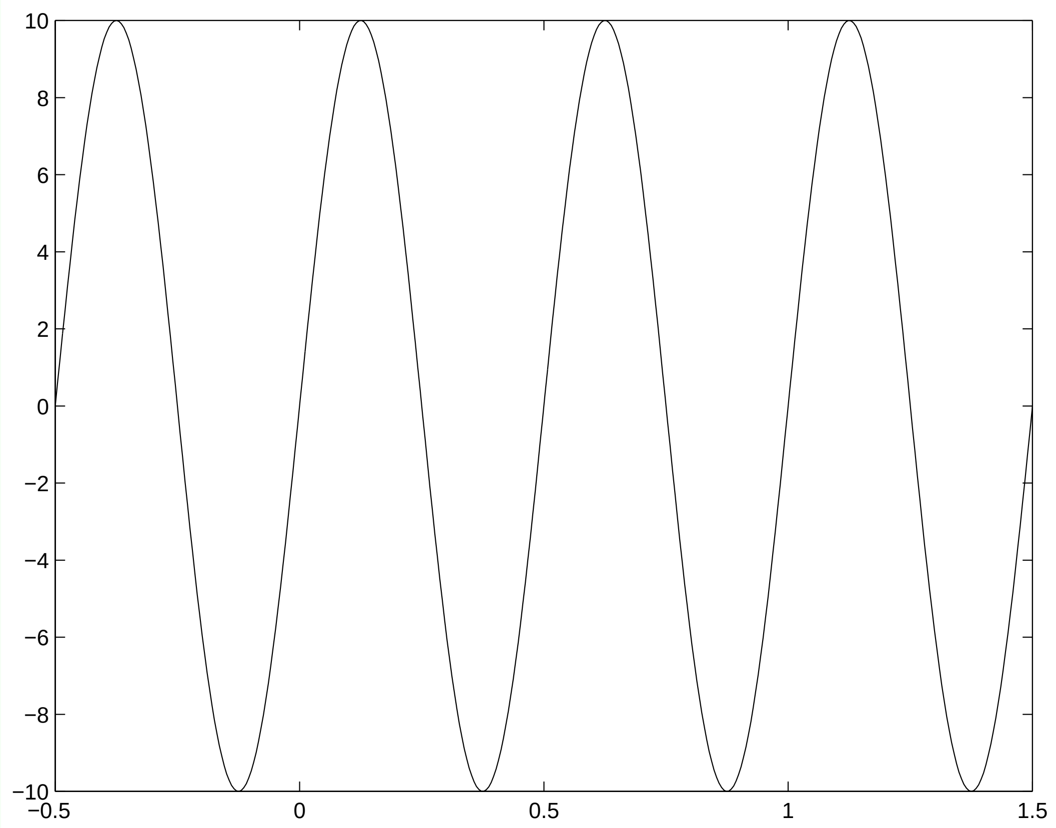
$x_+ = xI(x > 0)$, with $n = 200$, $\sigma_\epsilon^2 = 0.0015^2$, σ
$\mu_x = 0.5$ and $\sigma_x^2 = 0.25^2$.



Case 5 A modification of Case 4 of **CMR** so that

$$m(x) = 10\ \sin(4\pi x),$$

with $n = 500$, $\sigma_\epsilon^2 = 0.05^2$, $\sigma_u^2 = 0.141^2$, $\mu_x = 0.5$ and



Case 6 The same as Case 1 above except that X is chi-square(4) random variable. (Tests robustness tion of the structural assumptions.)

Mean Squared Bias $\times 10^2$					
Method	Case 1	Case 2	Case 3	Case 4	Ca
Naive	5.59	4.92	5.21	1,108	3,
Bayes	**0.78**	**0.38**	1.04	17.4	4
Structural, 5 knots	1.38	0.62	**0.46**	3.7	8
Structural, 15 knots	1.44	0.60	0.66	**3.3**	**2**

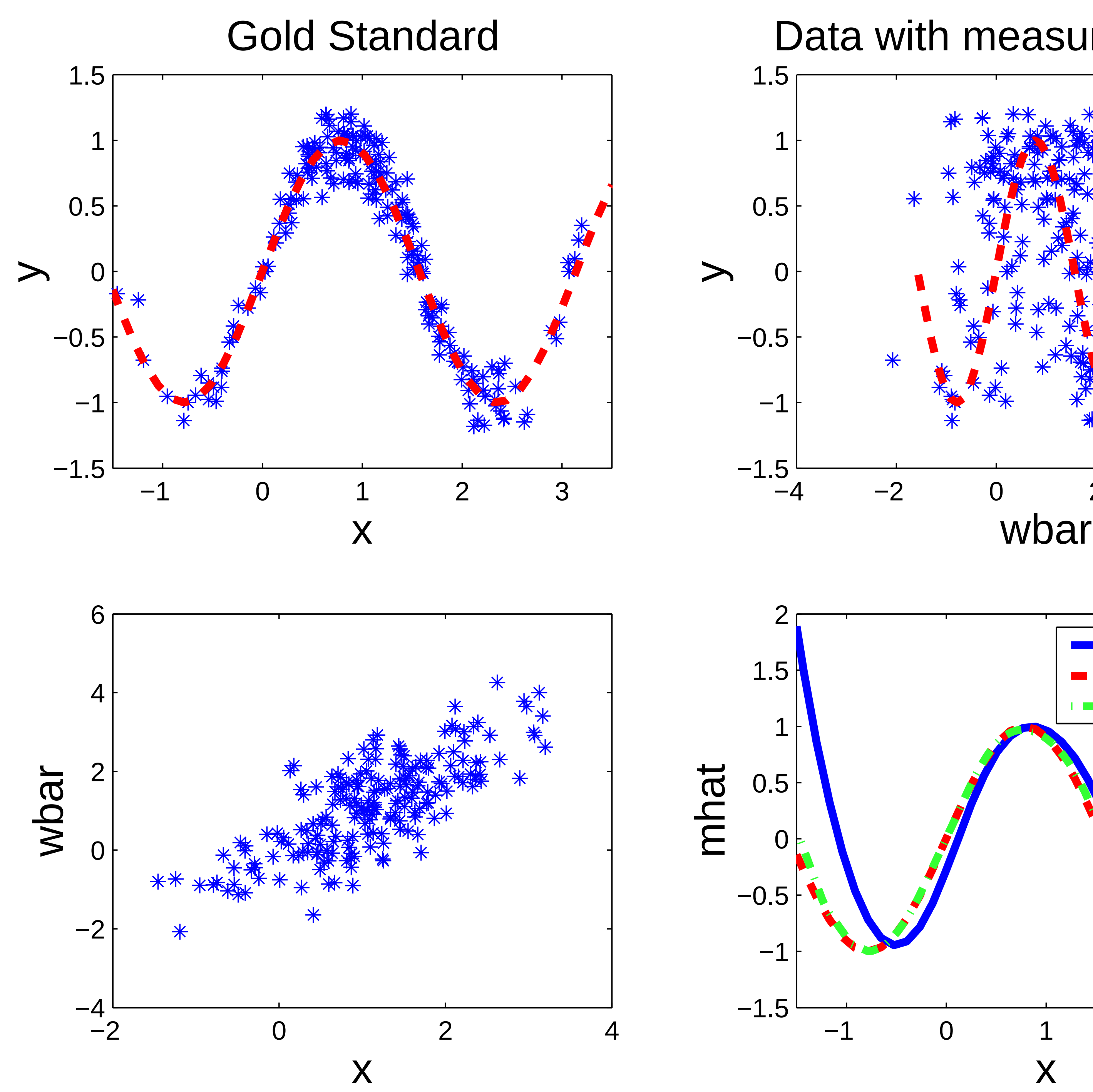
Mean Squared Error $\times 10^2$					
Method	Case 1	Case 2	Case 3	Case 4	Ca
Naive	6.91	5.57	5.38	1,155	3,
Bayes	**2.84**	**1.56**	**1.47**	**195**	1,
Structural, 5 knots	8.17	3.82	1.73	217	2,
Structural, 15 knots	9.90	5.40	1.85	237	**7**

Results based on 200 Monte Carlo simulations for each

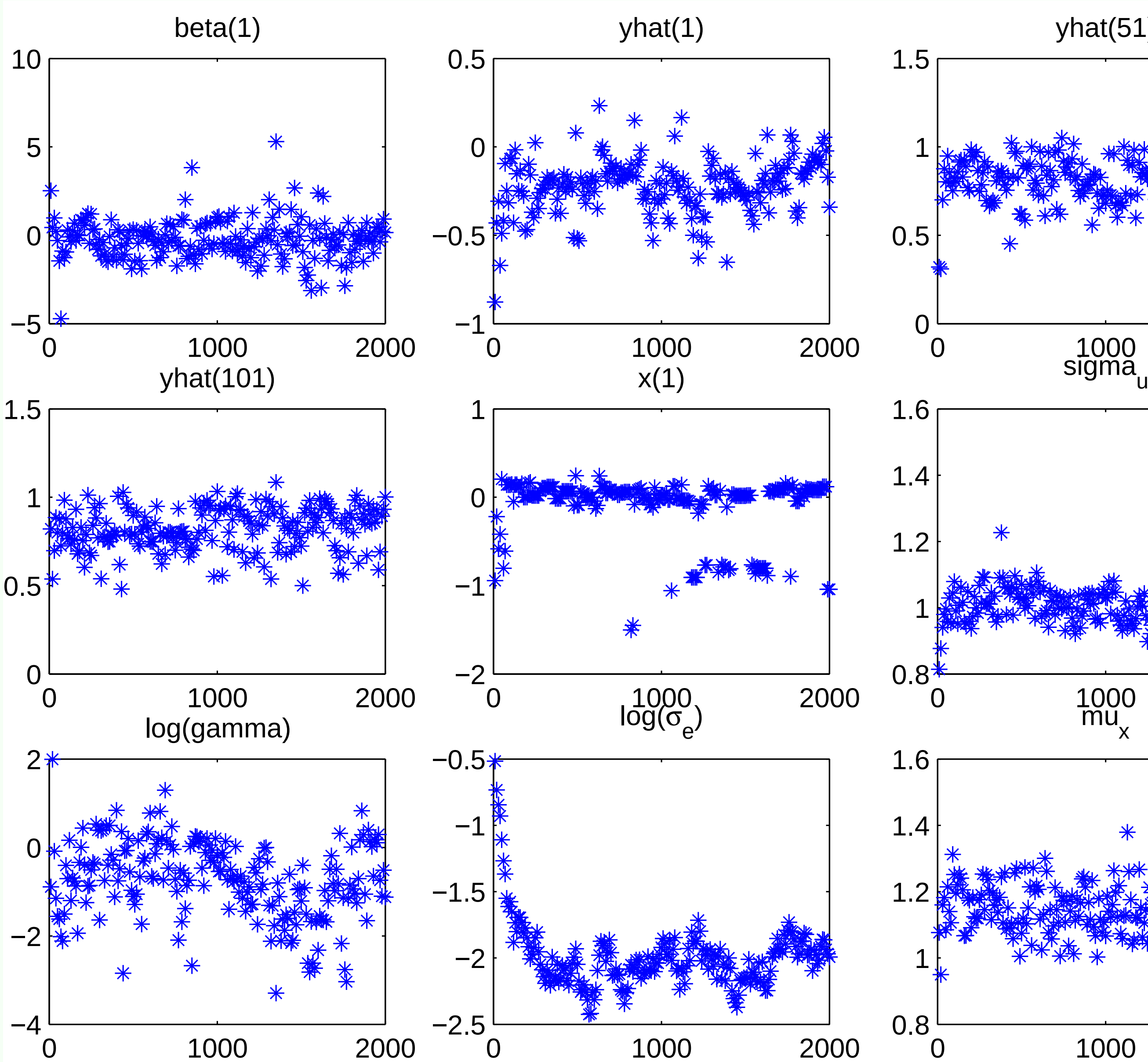
was not included in the table — it was not among the b

EXAMPLE — SIMULATED

- $Y = \sin(2X) + \epsilon$
- X is $N(1,1)$
- $\sigma_u = 1$
- $\sigma_e = 0.15$
- $n = 201$
- $n_i = 2$ for all i
- 15 knot quadratic P-splines
- 2,000 iterations of Gibbs. First 667 deleted as bur



Gold Standard
Data with measu
y
x
wbar
y
wbar
x
mhat
x



Results of Gibbs Sampling. Every twentieth ite

Note: $X(1) = -1.45$ **and** $\overline{W}(1) = -0.8$**. Also, log**$(\sigma$

EXAMPLE — SIMULATED

What does the Bayes approach work so well? Here's

tion:

Bayes uses all possible information to estimate

cially, $m(X)$.

- $\|m(X) - E\{m(X)|\mathbf{W}, \boldsymbol{Y}, \text{other param.}\}\|$
  $\approx \|m(X) - \text{ave}\{m(\widehat{X})\}\| = 2.47$
- $\|m(X) - m(E\{X|\mathbf{W}, \boldsymbol{Y}, \text{other param.}\})\|$
  $\approx \|m(X) - m(\text{ave}\{\widehat{X}\})\| = 4.67$
- $\|m(X) - m(E(X|\overline{W}))\| = 10.25$
- $\|m(X) - m(\overline{W})\| = 12.36$

DISCUSSION

- With the work of **CMR** and **BCR** we now have r
ficient estimators for nonparametric regression w
ment error.
  - SIMEX (LPR and splines) — in **CMR**
  - (Flexible) Structural splines — in **CMR**
  - Fully Bayesian (hardcore structural) — in **BCR**

DISCUSSION

- With **BCR** we have a methodology that
  - automatically selects the amount of smoothing
  - estimates the unknown X's
  - allows inference that takes account of the effects
    parameter selection and measurement error
- Most efficient methods appear to be structural, th
  may be competitive
  - hardcore structural methods seem reasonably r