NONPARAMETRIC REGRESSION AND MEASUREMENT ERROR

Raymond J. Carroll

Department of Statistics

Department of Biostatistics & Epidemiology

Faculty of Nutrition

Center for Environmental and Rural Health

Texas A&M University

David Ruppert

School of Operations Research and Inductrial Relations

Cornell University

AND CHECK OUT:

http://stat.tamu.edu/∼carroll **(most papers available, as well as some matlab software)**

http://stat.tamu.edu/B3NC **(Training program in Biology & Bioinformatics)**

OUTLINE

- Model
- Consistent estimation and rates of convergence
- Functional Method: SIMEX
- Structural methods
- Example
- Comments

THE BASIC MODEL

- Nonparametric regression:

$$Y = m(X) + \epsilon, \epsilon \sim (0, \sigma^2)$$

$m(\cdot)$ unspecified

- Measurement error: observe

$$W = X + U, U = \text{Normal}(0, \sigma_u^2) \tag{1}$$

  - Assume σ_u^2 known for today

- Model (1) is more general than it looks

- Since $m(\cdot)$ is unspecified, you need only need that (1) holds **after a transformation**

$$\begin{aligned} m(x) &= m[\exp\{\log(x)\}] \\ &= m_*\{\log(x)\}. \end{aligned}$$

- See work by **Nusser, et al.** and **Eckert & Carroll** for such transformations

THE BASIC MODEL

$$Y = m(X) + \epsilon, \epsilon \sim (0, \sigma^2)$$

$$W = X + U, U = \text{Normal}(0, \sigma_u^2)$$

- The issue broadly breaks down into what you must assume about the X's
- **Functional**: assume nothing about the X's
- **Structural**: assume a (possibly flexible) parametric distribution for the X's

DECONVOLUTION

$$Y = m(X) + \epsilon, \epsilon \sim (0, \sigma^2)$$

$$W = X + U, U = \text{Normal}(0, \sigma_u^2)$$

- Estimation of the regression function $m(\cdot)$ **without knowing the density of X**
  - Closely related to density deconvolution: estimate density of X from the W's alone
- Density deconvolution (with Stefanski & Hall, 1988) and nonparametric regression (Fan & Truong, 1993) are **nearly impossible**
- No globally consistent estimator of $m(x)$ can have a rate of convergence faster than $\log(n)$
  - $\log(10,000,000) \approx 16$
  - A sample of size **16** in parametric regression has approximately the same information as a sample of size **10,000,000** for consistent nonparametric regression

DECONVOLUTION

- The deconvolution results mean that if you insist on being consistent nonparametrically
    - and if you insist on no assumptions about the density of the latent X
    - **The you cannot get a decent estimator**
- Fan & Truong constructed an ”estimator” using our deconvoluting kernel
    - **It’s awful**
    - Bandwidth selection essentially impossible
    - Even cheating and selecting the bandwidth that minimizes MSE, the method has **MSE’s an order of magnitude larger** than our other methods
    - Theory really means something here

WHAT CAN WE DO??

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

- One method is to allow a **little bit** of inconsistency
  - Still make no assumptions about the density of X
  - But construct an estimator which has less bias than the naive estimator that ignores error
- The naive estimator has a bias of order $O_p(\sigma_u^2)$
- The **SIMEX estimator** of Cook and Stefanski has a bias of order $O_p(\sigma_u^6)$
- SIMEX is **general purpose**: kernels, splines, etc.

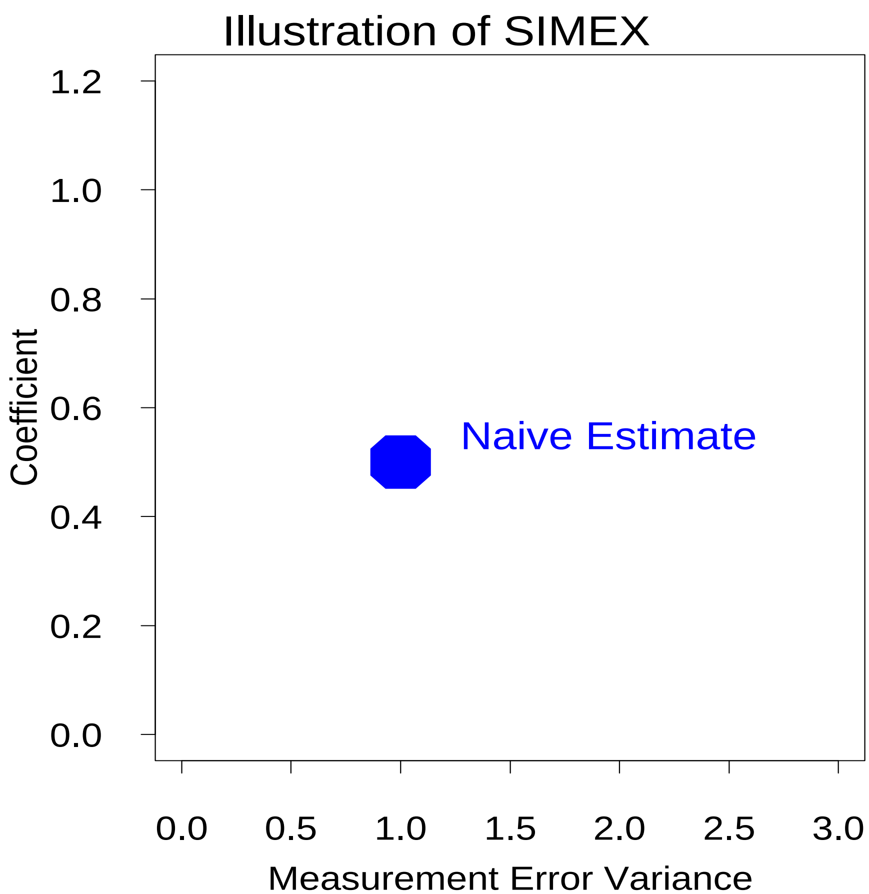


Illustration of SIMEX
Coefficient
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Naive Estimate
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Measurement Error Variance

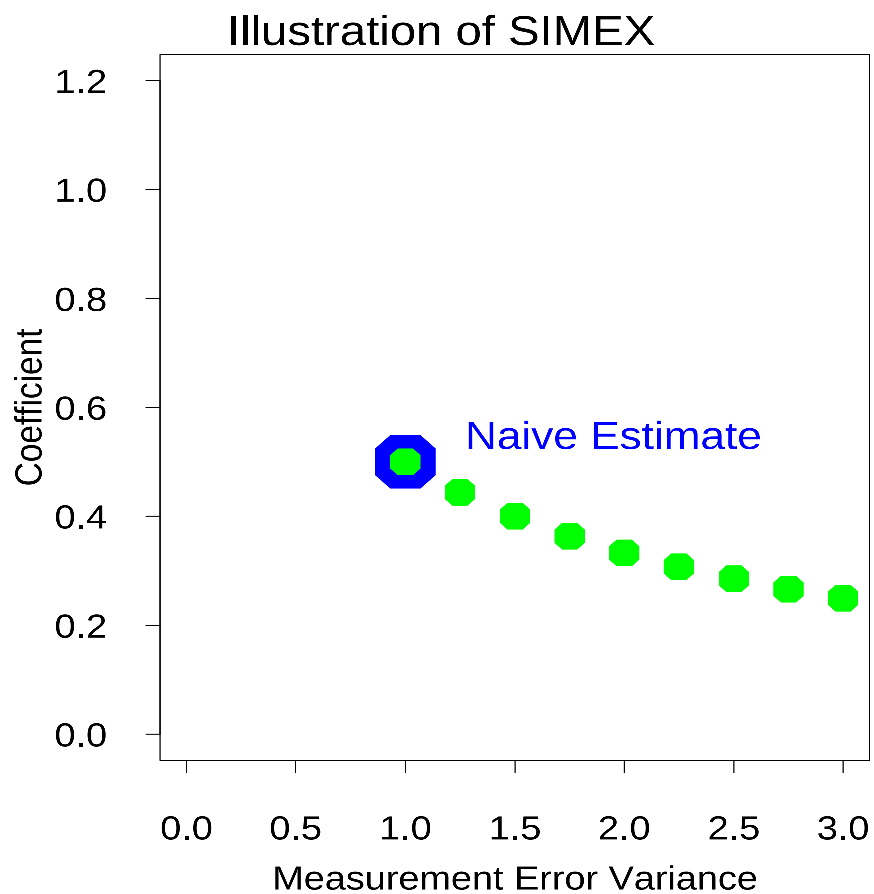


Illustration of SIMEX
Naive Estimate
Coefficient
Measurement Error Variance
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0

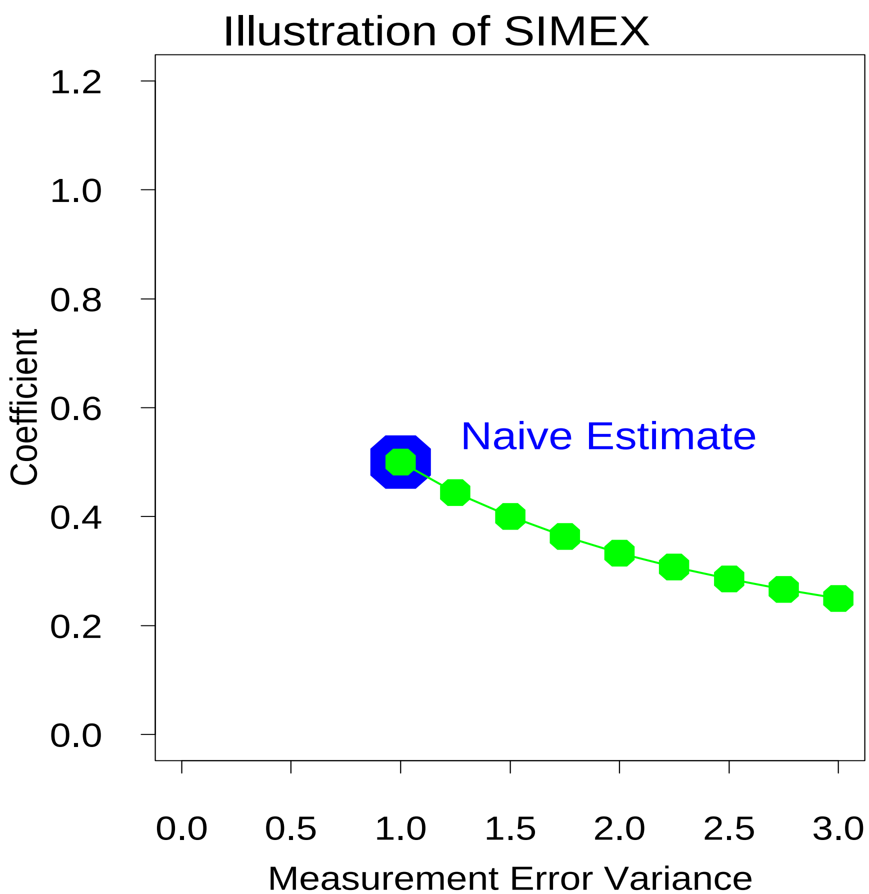


Illustration of SIMEX
Coefficient
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Naive Estimate
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Measurement Error Variance

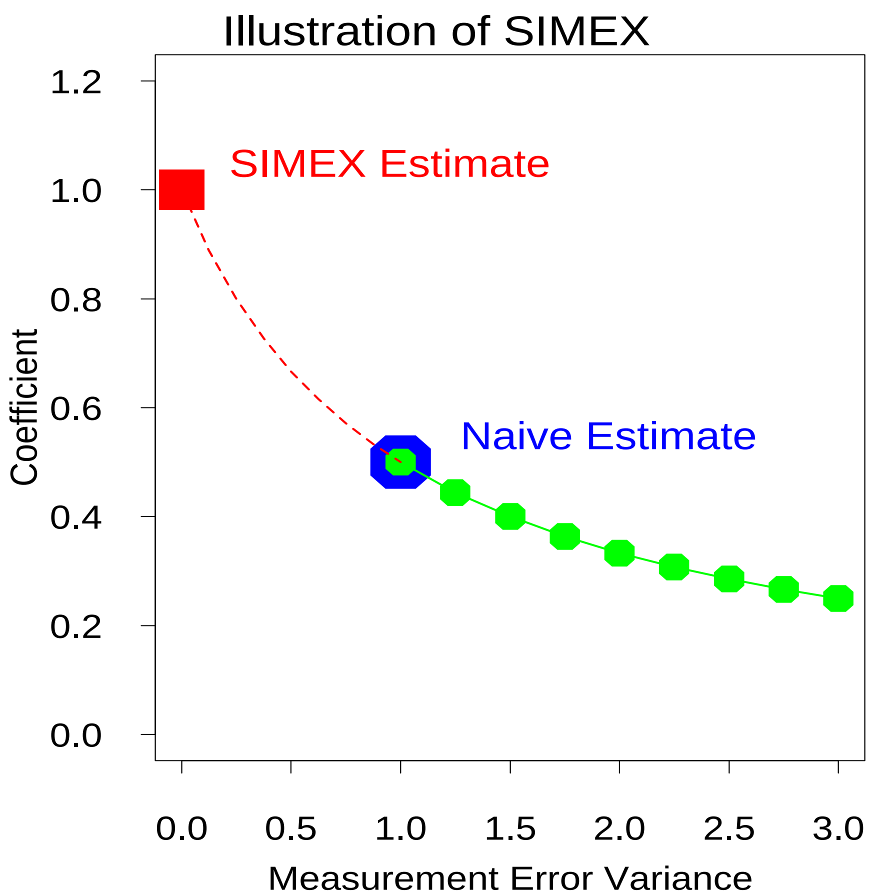


Illustration of SIMEX
SIMEX Estimate
Naive Estimate
Coefficient
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Measurement Error Variance

SIMEX

- We have derived the asymptotics for SIMEX in kernel regression
  - Bias is $O_p(\sigma_u^6)$
  - Rate of convergence is standard $n^{-2/5}$
  - **Approximately consistent** estimation with standard rates
  - Deconvolution is **globally consistent** with slow rates
  - SIMEX has vastly beaten deconvolution in all our simulations
  - Asymptotic variance is the same as if measurement error was ignored, but multipled by a factor depending only on the extrapolant function
  - linear–quadratic–cubic ratios are 1–9–52

SPLINES

$Y = m(X) + \epsilon, W = X + U$

- Write $m(\cdot)$ in terms of **basis functions**:

$$m(x) = \sum_{j=1}^{M} \theta_j B_j(x)$$

  - **Smoothing splines**
  - **P–splines** (Eiler & Marx, with Ruppert, like Splus), e.g.,

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \sum_{j=3}^{M} \theta_j (x - \xi_j)_+^2$$

- P–splines and smoothing splines are **”typically” equivalent**, even for $M = 15$, at least in our context
- We place knots at quantiles of the W–distribution

MEAN–BASED STRUCTURAL METHODS

$$m(x) = \Sigma_{j=1}^{M} \theta_j B_j(x)$$

- It follows that

$$E(Y|W) = \Sigma_{j=1}^{M} \theta_j E\{B_j(X)|W\} = \Sigma_{j=1}^{M} \theta_j Q_j(W)$$

- If X is parametric, and U is normal, then $[X|W]$ is known and $Q_j(\cdot)$ is readily calculated
- **Penalization**: for some matrix D, minimize

$$\Sigma_{i=1}^{n}\{Y_i - \Sigma_{j=1}^{M} \theta_j Q_j(W)\}^2 + \alpha \underline{\theta}^T D \underline{\theta}$$

- **Ridge–type estimator**: α must be estimated
  - Shrinkage of B_j to $W \Longrightarrow Q_j$ **nearly singular**
  - $\alpha = 0$ yields **enormously variable estimates**
  - We developed a **"minimize estimated MSE"** shrinkage method

SPLINES: FULLY STRUCTURAL

- If we specify a distribution for X and ϵ, then we have a **fully specified parametric model** using the standard mixed model formulation:

$$\begin{aligned} Y &= \theta_0 + \theta_1 x + \theta_2 x^2 + \sum_{j=3}^{M} \theta_j (x - \xi_j)_+^2 + \epsilon \\ \theta_j &= \text{Normal}(0, \sigma_\theta^2) \text{ for } j \geq 3 \\ \epsilon &= \text{Normal}(0, \sigma_\epsilon^2) \\ X &= \text{Normal}(\mu_x, \sigma_x^2) \\ W &= \text{Normal}(X, \sigma_u^2) \end{aligned}$$

- Good starting values are easy: previous method or SIMEX
- A penalty is incorporated naturally through the variance component σ_θ^2
- Computed via MCMC: program on my web site

COMMENTS ON FULLY STRUCTURAL

- We have investigated allowing X to have a flexible distribution, i.e., mixture of ≤ 3 normals
- More complex computationally
- Useful for severely skew X, not much gain otherwise (or loss!)
- Still, **structural spline is remarkably robust to deviations from normality for X**

SPECIAL CASE: SIMULATION

- $n = 200$ with

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

- X standard normal, 30% of observed variability in W due to error
- Frequentist MSE
  - Naive smoothing spline: 5.6
  - SIMEX smoothing spline: 4.7
  - Mean–based P–spline, $M = 15$: 3.8
  - Bayes P–spline, $M = 40$: 1.6
- Note: **Bayes method is a better frequentist estimator**
- Here are some examples for $\sin(2x)$ with different amounts of measurement error

SIMEX

- As I will show via simulated data, the naive estimate in the presence of measurement error **loses all features**
- That is, with sufficiently large measurement error, the Y and W data look like a polynomial
- The SIMEX method can be expected to perform fairly poorly here.
- SIMEX adds error to a problem that has already lost its features
- Each SIMEX step is going to fit a polynomial
- Thus, SIMEX with sufficiently large measurement error is typically going to be a corrected polynomial fit.

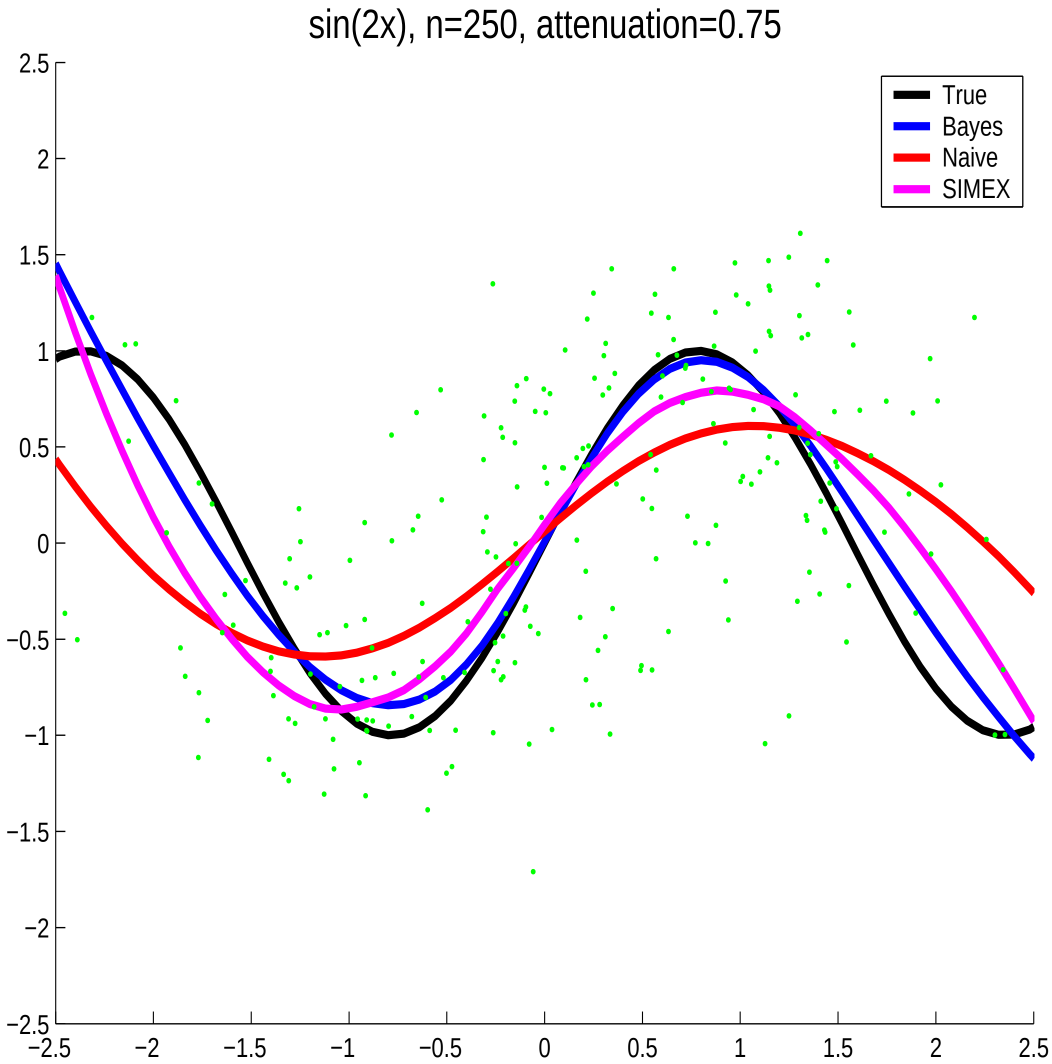


Figure 1: Note how the naive fit captures some of the features, so SIMEX does much better

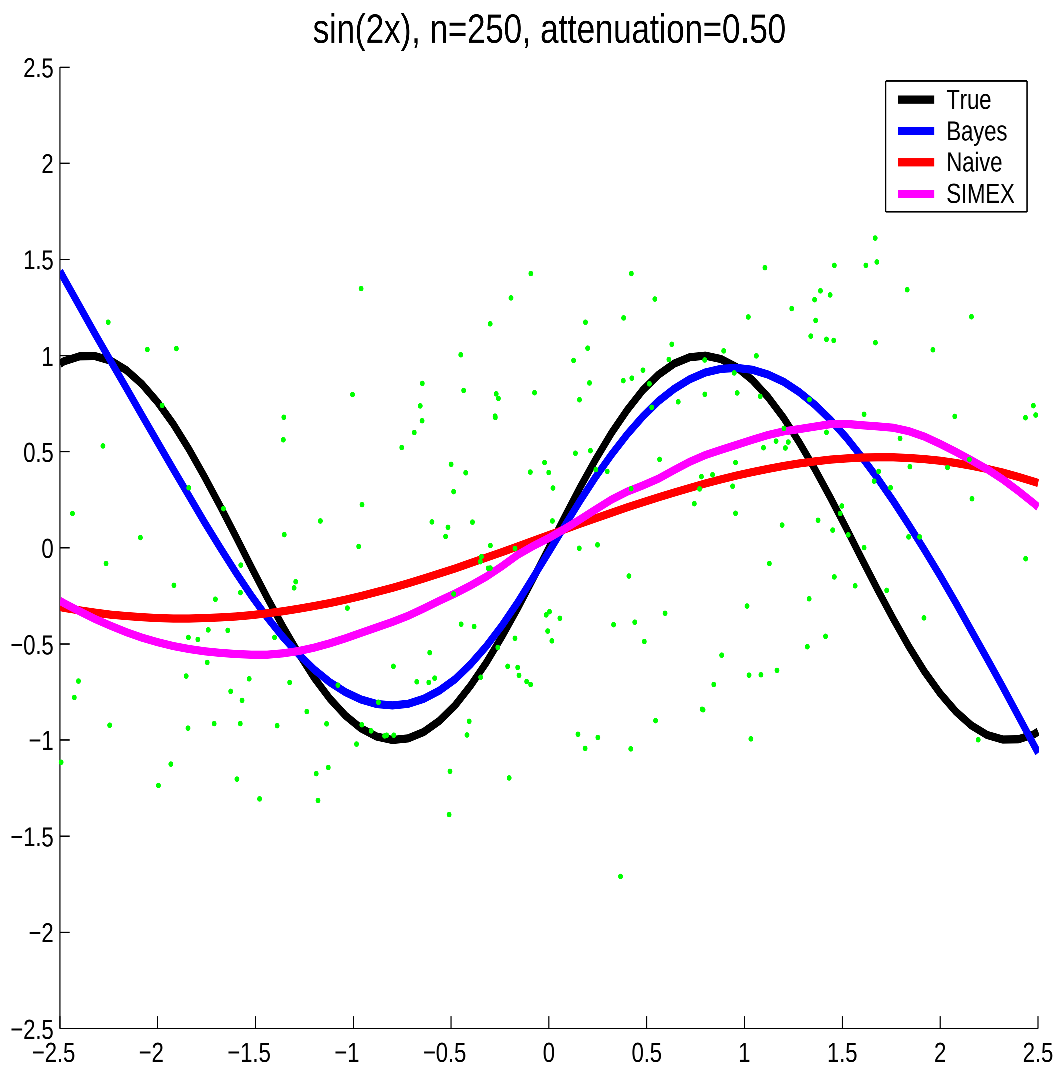
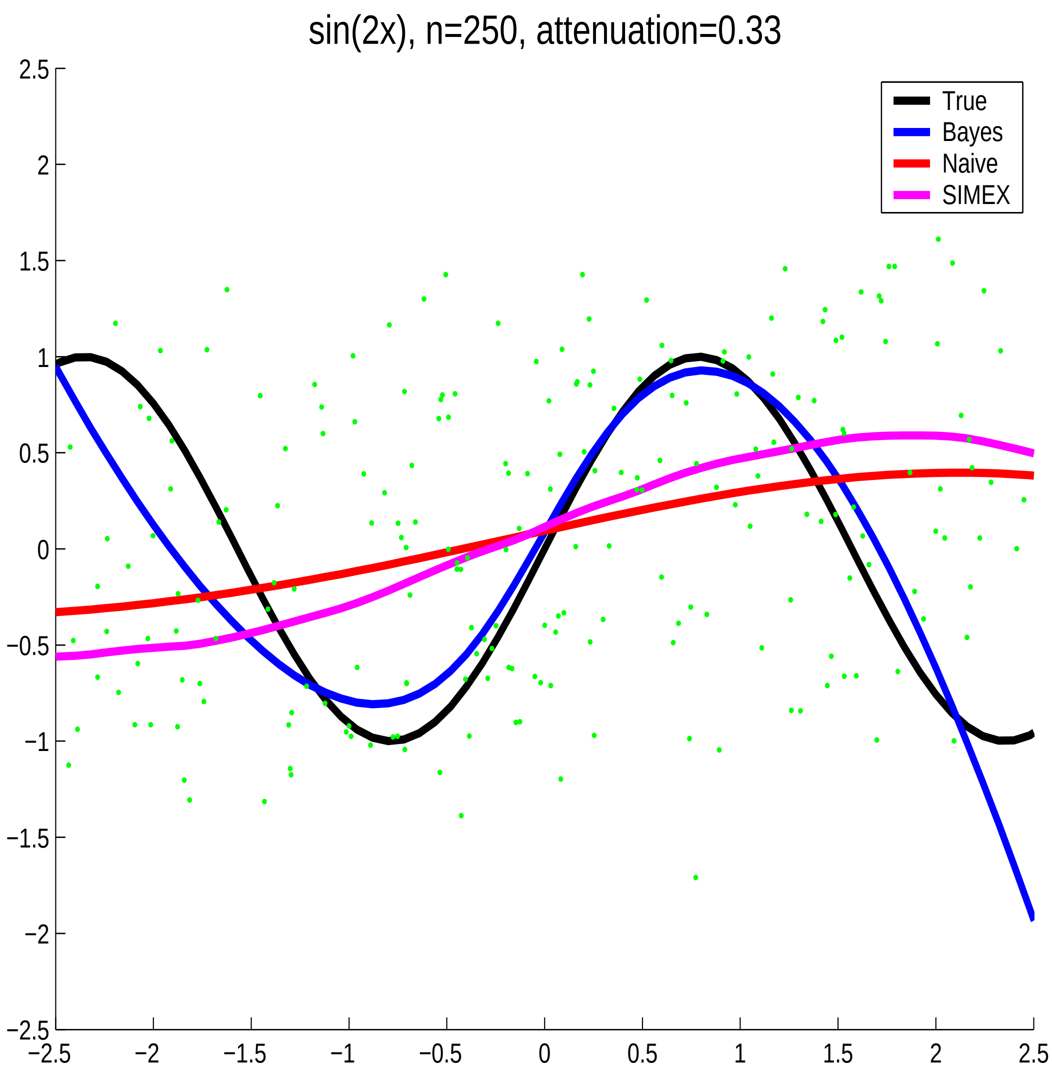


Figure 2: Note how the naive fit captures **NONE** of the features, so SIMEX does **NO** better


sin(2x), n=250, attenuation=0.33
True
Bayes
Naive
SIMEX
2.5
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2.5
−2.5
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
2.5

WHY DOES THE STRUCTURAL SPLINE WORK SO WELL?

- Take $m(x) = \sin(2x)$, $X \sim \text{Normal}(1,1)$, $\lambda = 2/3$, $\sigma_\epsilon^2 = 0.15$.
- Estimate X
  - Estimate #1: $E(X|W)$
  - Estimate #2: Mean of X over MCMC samples
- Ratio of MSE for $\sin(2\widehat{X}) \approx 3$
- Structural spline with MCMC gives a **much** better estimate of X
- Structural Spline with MCMC uses the assumed distribution of Y given X in a strong way
- Suggests **potential** for model non–robustness

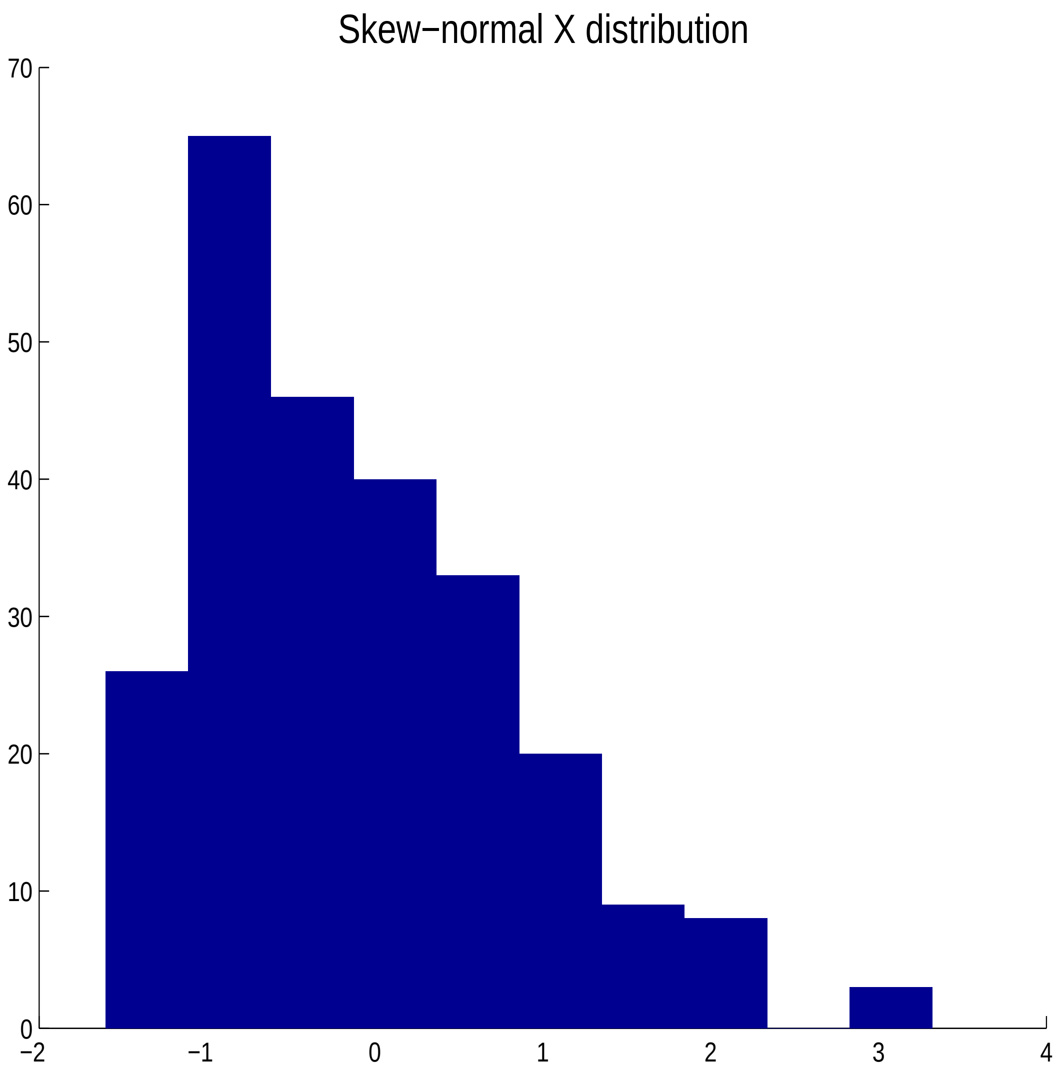


Figure 3: This is the distribution for X and ϵ, which are both **assumed** to be normal

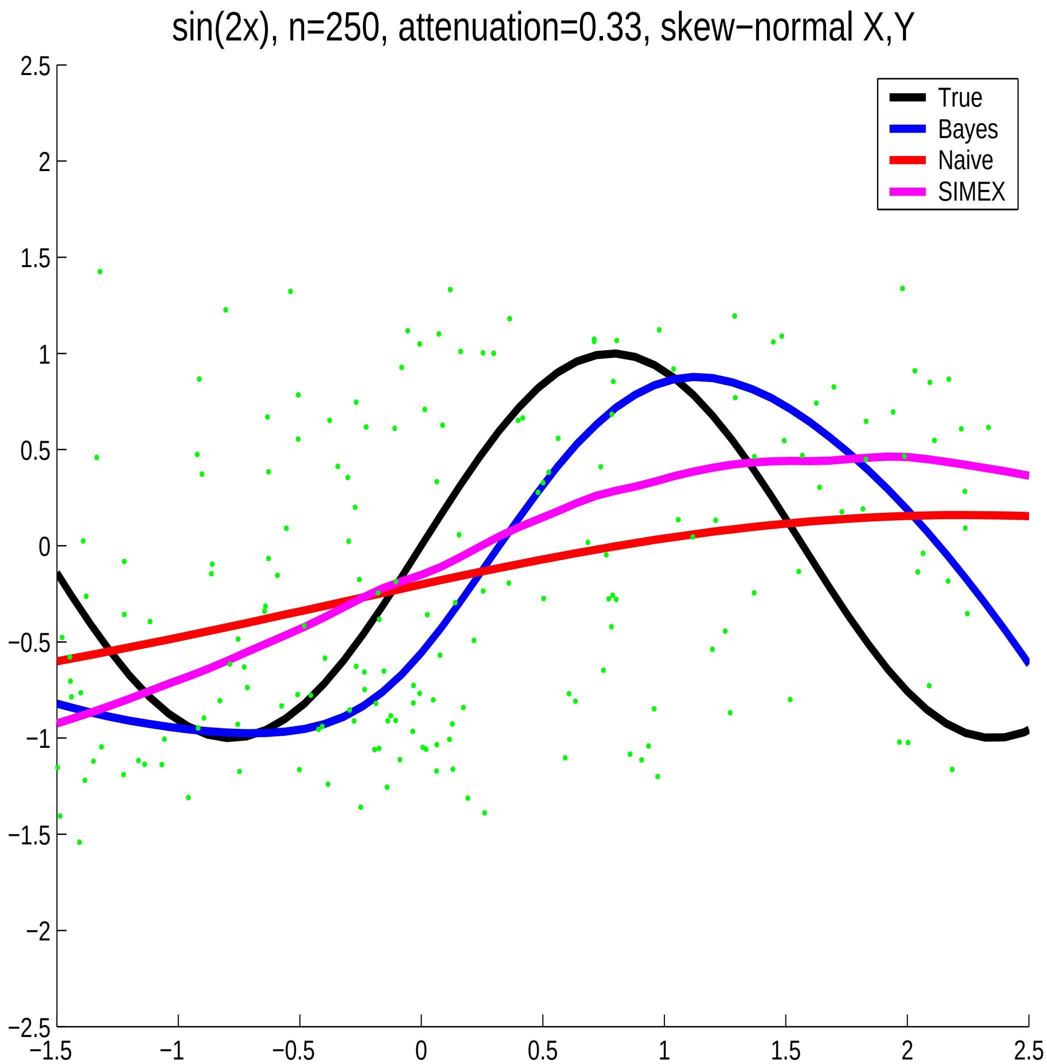


Figure 4: Here you assume that X and ϵ are normal, when they are skew–normal. Note how the Bayes fit is still much better than naive or SIMEX, but clearly worse than the normal X and ϵ simulation

SUMMARY OF METHODS

- **Naive**: great estimate of the wrong thing
- **Deconvolution**: lousy estimate of the right thing
- **SIMEX**: approximately consistent, no assumptions, **cannot cope with large measurement error**
- **Mean function** P–splines
  - Parametric model for [X]
  - Flexible mean function
  - **Ad hoc smoothing**
- **Structural splines**
  - Parametric model for Y given X, [X]
  - MCMC incorporates penalty naturally
  - Results great even when $X \sim \chi^2(4)$
  - **Few** results yet when Y given X is mismodeled

CONCLUSIONS

- The more (correct!) assumptions you make, the more efficient the estimator
- Heirarchy
  - Global consistency, [X], Y given X unspecified
  - Approximate consistentcy, [X], Y given X unspecified
  - Flexible means, [X] specified
  - Bayes, [X], Y given X specified
- **It is remarkable** how well you can extract the features of the problem in the presence of measurement error