Modeling Nonlinear Relationships among Latent Variables via Mixtures of Linear Structural Equations

Daniel Bauer, Ruth Mathiowetz,
Nisha Gottfredson, Jolynn Pek, Diane Losardo

The University of North Carolina at Chapel Hill
Supported by NSF SES-0716555

Modeling Nonlinear Effects

- Much attention given to modeling nonlinear effects in SEMs since Kenny & Judd (1984)
  - Product-indicant approaches
  - Method of moment approaches
  - ML or QML approaches
  - Bayesian approaches
- Focus has been on modeling nonlinear effects via known, low-order polynomial functions
- What if functional form is unknown?

A Semiparametric Approach

- Bauer (2005) proposed a semiparametric modeling approach
  - Fit a mixture of linear structural equation models
  - Aggregate over components of the mixture to obtain smooth approximation of nonlinear function
  - Model is locally linear (within mixing component), but globally nonlinear (across mixing components)

The SEM

The SEM (y-side form) consists of…

measurement model:

$$\mathbf{y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i; \quad \boldsymbol{\varepsilon}_i \sim N_p(\mathbf{0}, \boldsymbol{\Theta})$$

linear structural model:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i; \quad \boldsymbol{\zeta}_i \sim N_q(\mathbf{0}, \boldsymbol{\Psi})$$

The SEM

- By implication, marginal distribution of $\mathbf{y}$ is normal

$$\phi[\mathbf{y};\ \boldsymbol{\mu}(\boldsymbol{\theta}),\ \boldsymbol{\Sigma}(\boldsymbol{\theta})]$$

with implied mean and covariance structure

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\alpha}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})^{-1'}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

Finite Mixture SEM

- Finite mixture SEM involves fitting SEMs within each of K latent components

$$f(\mathbf{y}) = \sum_{k=1}^{K} P(k)\phi_k\left[\mathbf{y};\ \boldsymbol{\mu}_k(\boldsymbol{\theta}_k), \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)\right]$$

- $P(k)$ are mixing probabilities
- K usually not estimated; chosen through model comparison
- Parameters can vary over components or be constrained to equality

Mixture Applications

- Direct:
  - Goal is to separate the population into hidden subgroups
  - Component distributions thought to reflect some physical reality (types of people, disorders, developmental pathways)
  - Interpretation focuses on component distribution parameters

Mixture Applications

- Direct:
  - Goal is to separate the population into hidden subgroups
  - Component distributions thought to reflect some physical reality (types of people, disorders, developmental pathways)
  - Interpretation focuses on component distribution parameters
- Indirect:
  - Goal is to use mixture to capture data features not easily recovered using other distributions
  - Component distributions are only a statistical expedience
  - Interpretation focuses on the aggregate mixture distribution

Nonlinear Effect Modeling by FM-SEM

- Locally linear within component:

$$E_k(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) = \boldsymbol{\alpha}_{2k} + \mathbf{B}_{21k}\boldsymbol{\eta}_1$$

Nonlinear Effect Modeling by FM-SEM

- Locally linear within component:

$$E_k(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) = \boldsymbol{\alpha}_{2k} + \mathbf{B}_{21k}\boldsymbol{\eta}_1$$

- Global function is nonlinear:

$$E(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) = \sum_{k=1}^{K} P(k \mid \boldsymbol{\eta}_1) E_k(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1)$$

Nonlinear Effect Modeling by FM-SEM

- Locally linear within component:

$$E_k(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) = \boldsymbol{\alpha}_{2k} + \mathbf{B}_{21k}\boldsymbol{\eta}_1$$

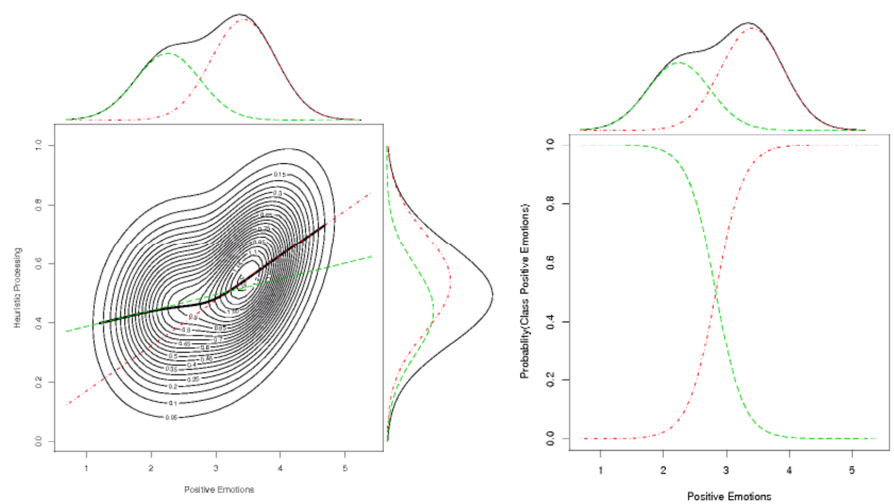
- Global function is nonlinear:

$$E(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) = \sum_{k=1}^{K} P(k \mid \boldsymbol{\eta}_1) E_k(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1)$$

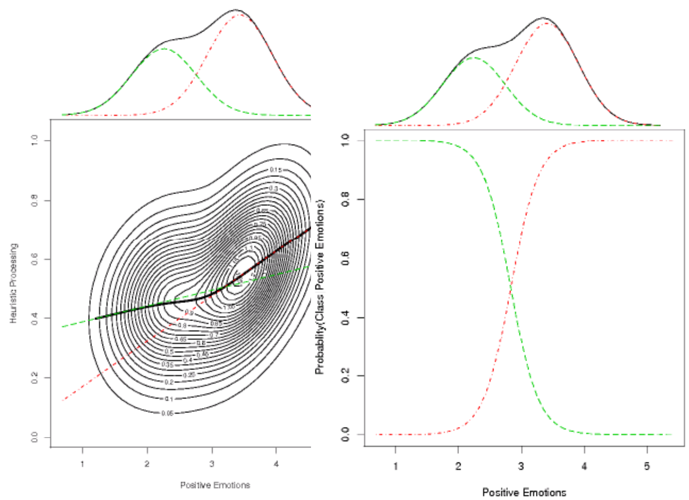
- Smoothing weights are conditional probabilities:

$$P(k \mid \boldsymbol{\eta}_1) = \frac{P(k)\phi_k(\boldsymbol{\eta}_1; \boldsymbol{\alpha}_{1k}, \boldsymbol{\Psi}_{11k})}{\sum_{k=1}^{K} P(k)\phi_k(\boldsymbol{\eta}_1; \boldsymbol{\alpha}_{1k}, \boldsymbol{\Psi}_{11k})}$$

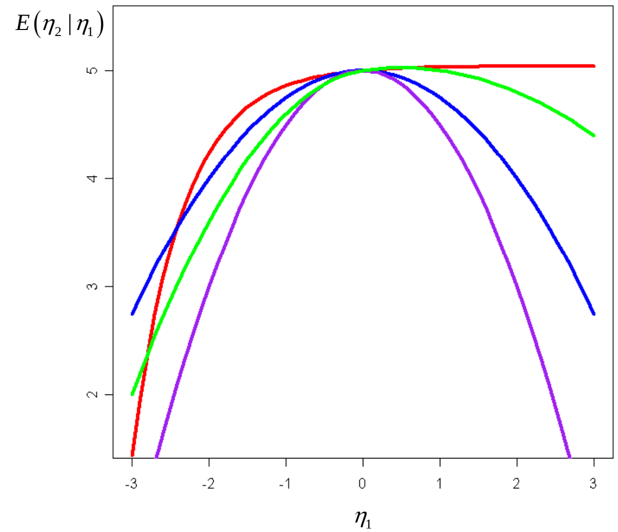
Bivariate Example



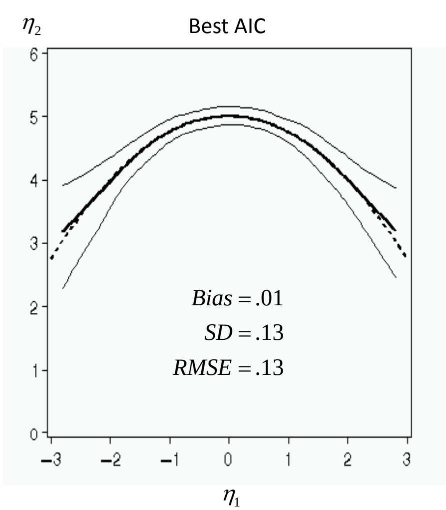
Bivariate Example

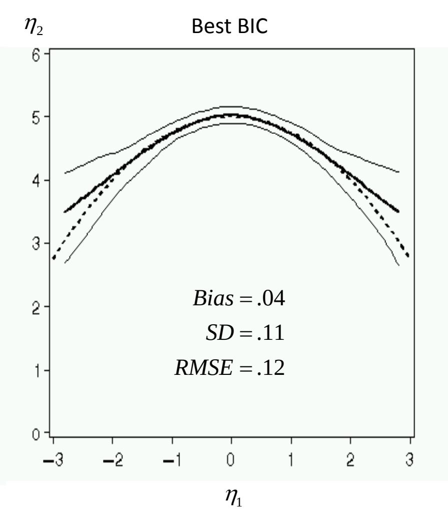


Does it Work?

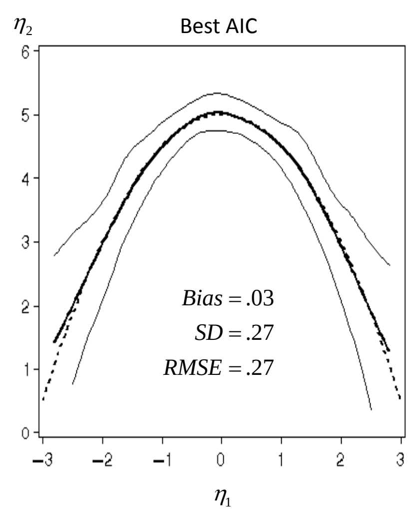


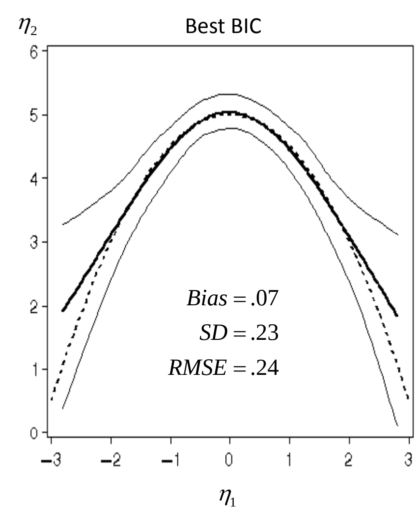
Small Quadratic



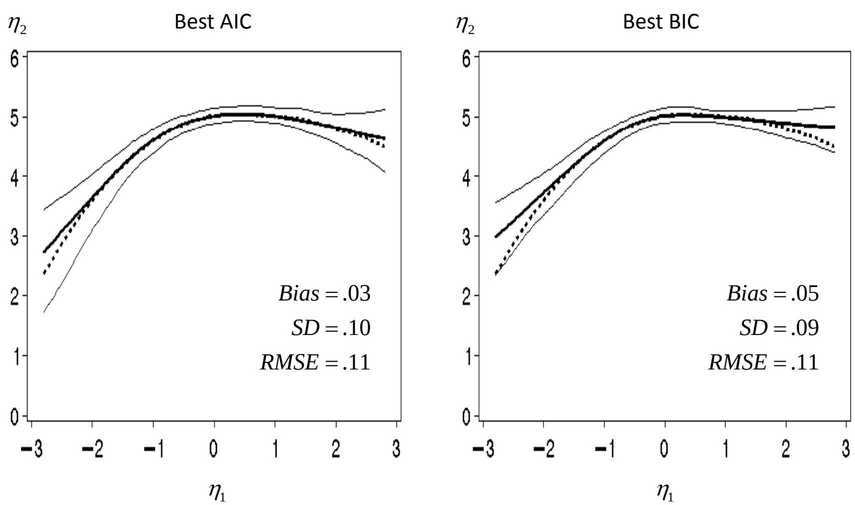


Large Quadratic

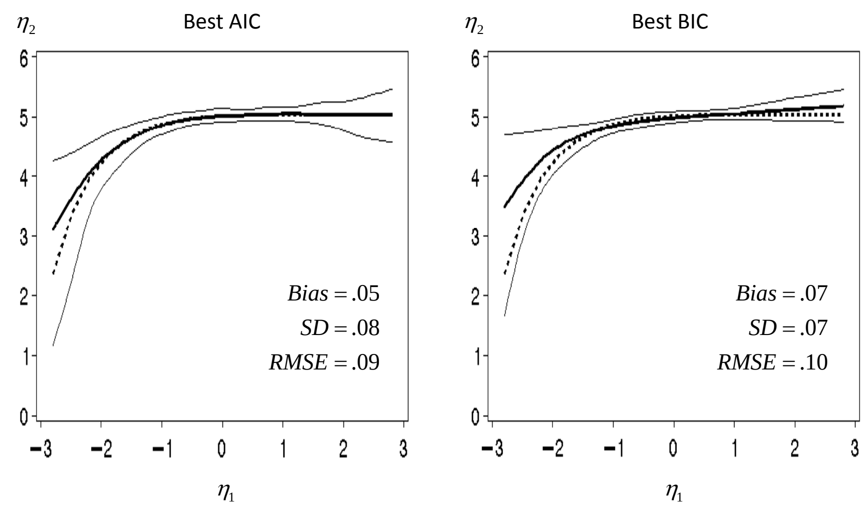




Quadratic Spline



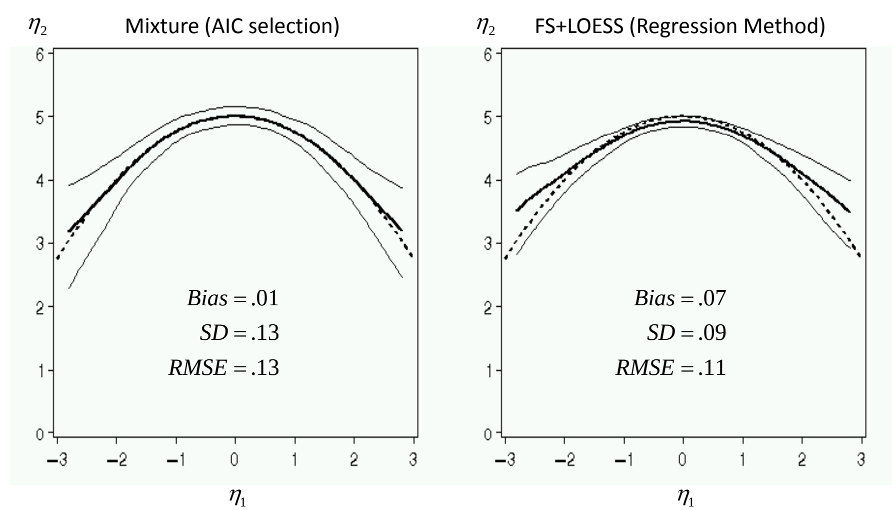
Exponential



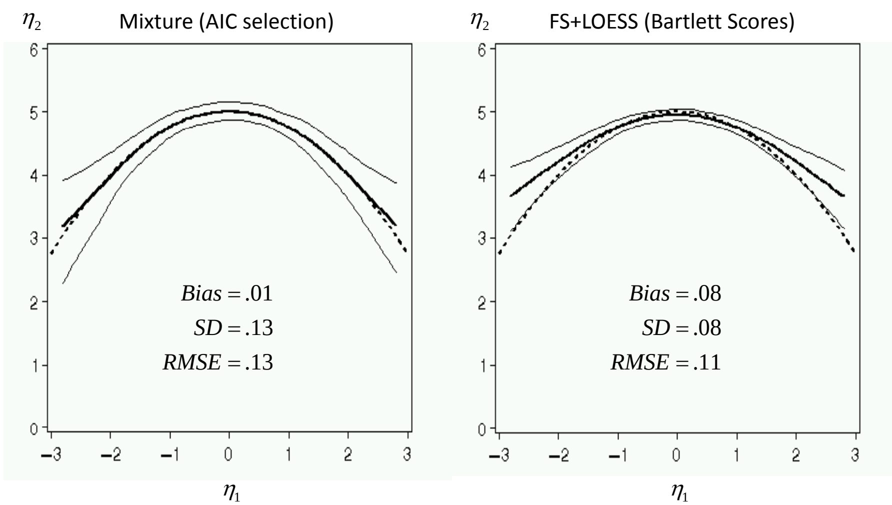
Is it Worth the Trouble?

- Why not just plot factor score estimates instead?
- Compared the FM-SEM approach to a three-step approach (FS+LOESS):
  - Fit standard linear SEM model
  - Generate factor score estimates for latent predictor and outcome
  - Perform LOESS regression on factor score estimates
- Anticipated less bias for FM-SEM

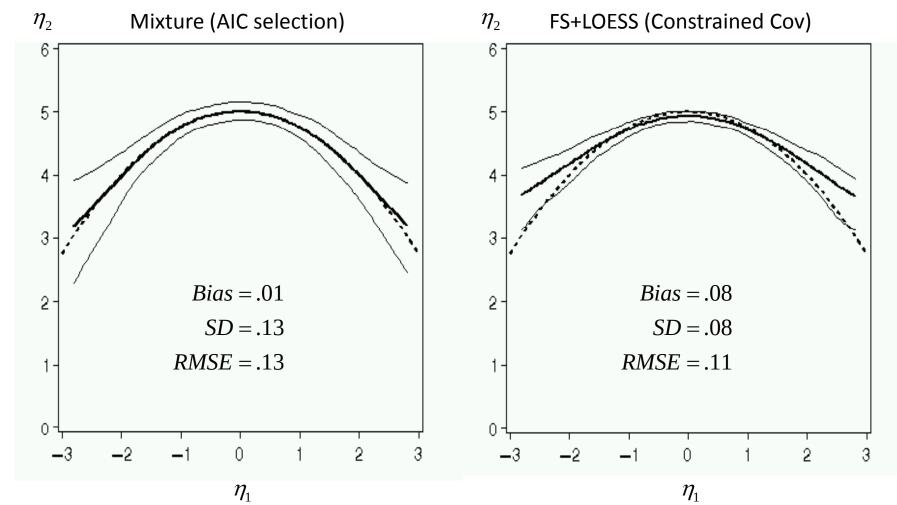
Small Quadratic



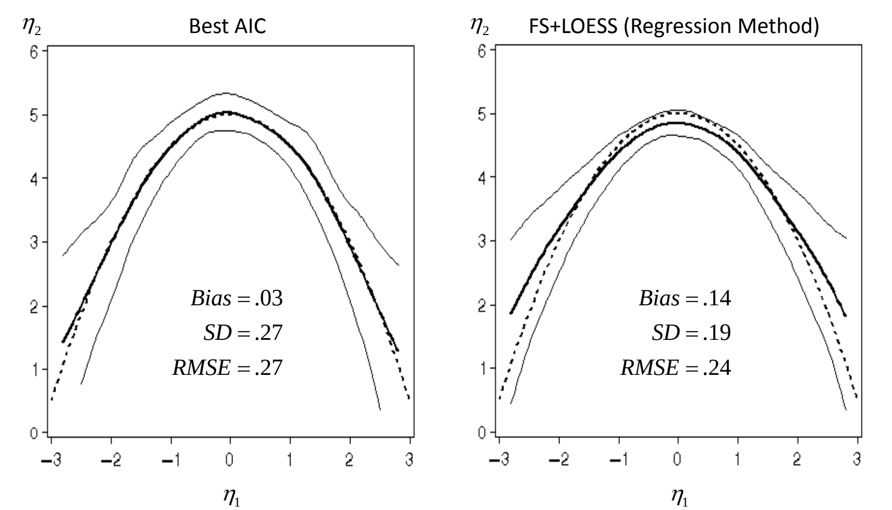
Small Quadratic



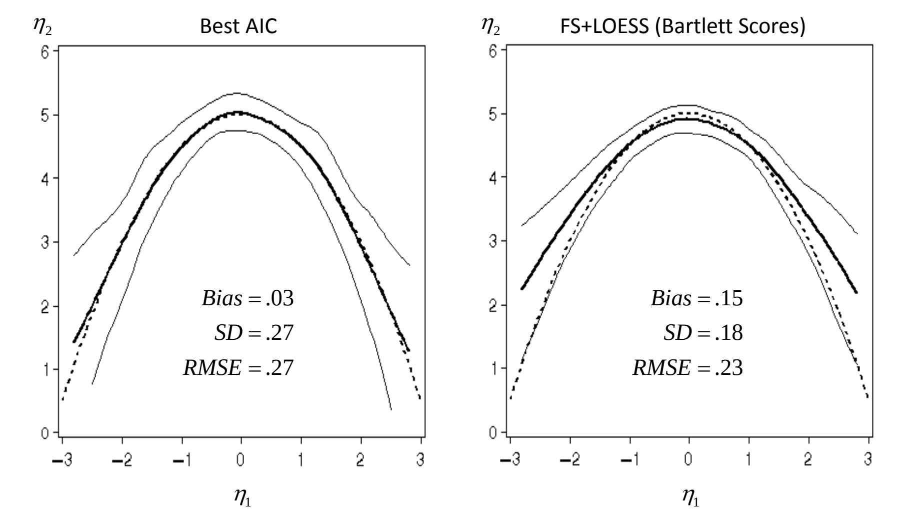
Small Quadratic



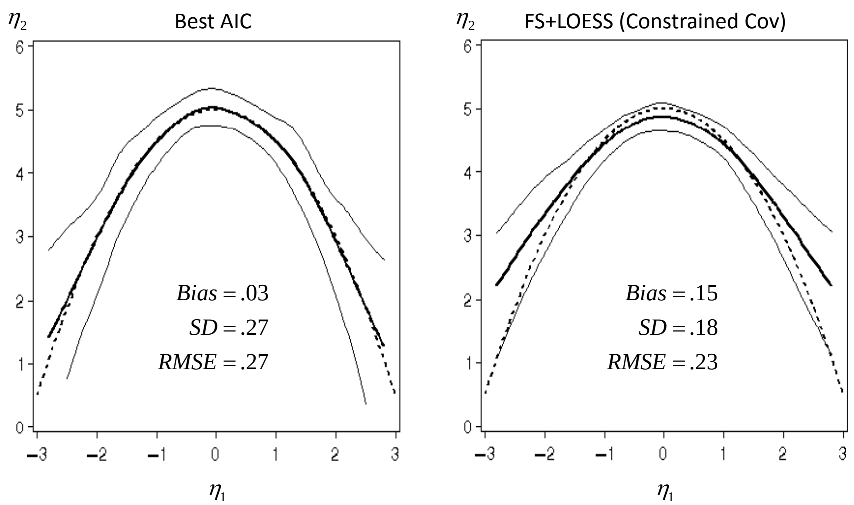
Large Quadratic



Large Quadratic



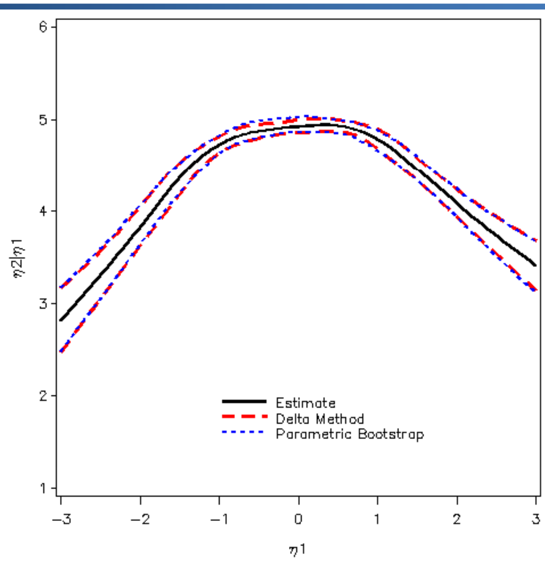
Large Quadratic



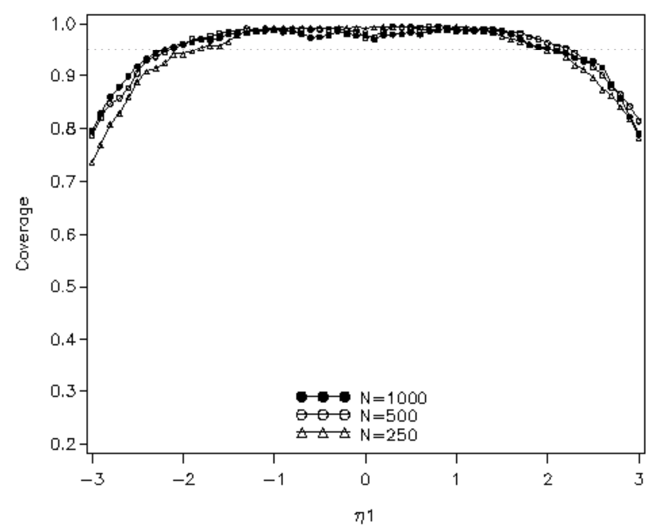
Can We Make Inferences?

- Since no *parameters* are estimated, inference must focus on *function*
- Monte Carlo plots have bands enclosing 95% of estimates across replications
- Similarly, want 95% confidence bands for any single replication
- Pointwise (non-simultaneous) confidence intervals constructed using two methods:
  - Delta Method
  - Parametric Bootstrap

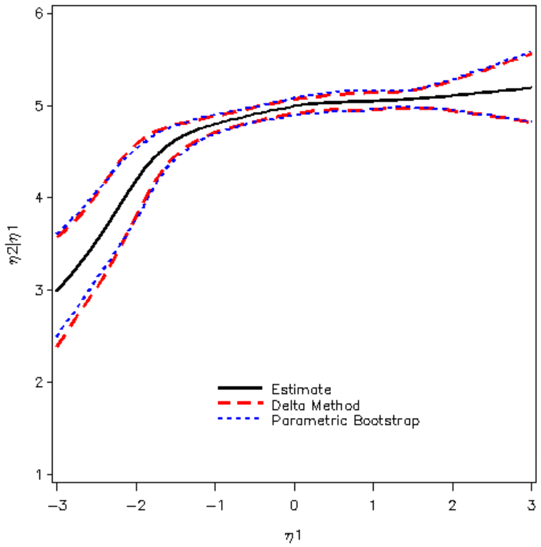
One Replication: Small Quadratic



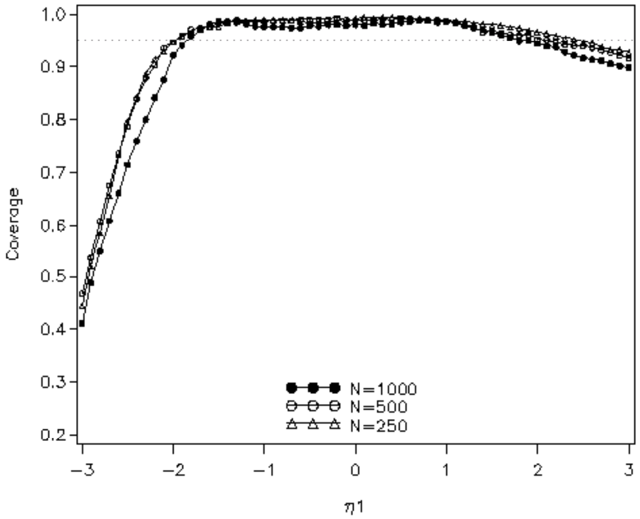
Coverage: Small Quadratic



One Replication: Exponential



Coverage: Exponential



Conclusions

- *Does it work?*
  - Yes, recovers function without a priori specification

Conclusions

- *Does it work?*
  - Yes, recovers function without a priori specification
- *Is it worth the effort?*
  - Unbiased over most of the range of the data
  - Bias and efficiency tradeoff

Conclusions

- *Does it work?*
  - Yes, recovers function without a priori specification
- *Is it worth the effort?*
  - Unbiased over most of the range of the data
  - Bias and efficiency tradeoff
- *Can we make inferences?*
  - Yes, with pointwise CIs, although coverage poor in tails and a bit high in the middle.

Conclusions

- *Does it work?*
  - Yes, recovers function without a priori specification
- *Is it worth the effort?*
  - Unbiased over most of the range of the data
  - Bias and efficiency tradeoff
- *Can we make inferences?*
  - Yes, with pointwise CIs, although coverage poor in tails and a bit high in the middle.
- *Where to next?*
  - Modeling nonlinear interactions for multiple latent predictors