Astrostatistics: Fri 07 Feb 2020

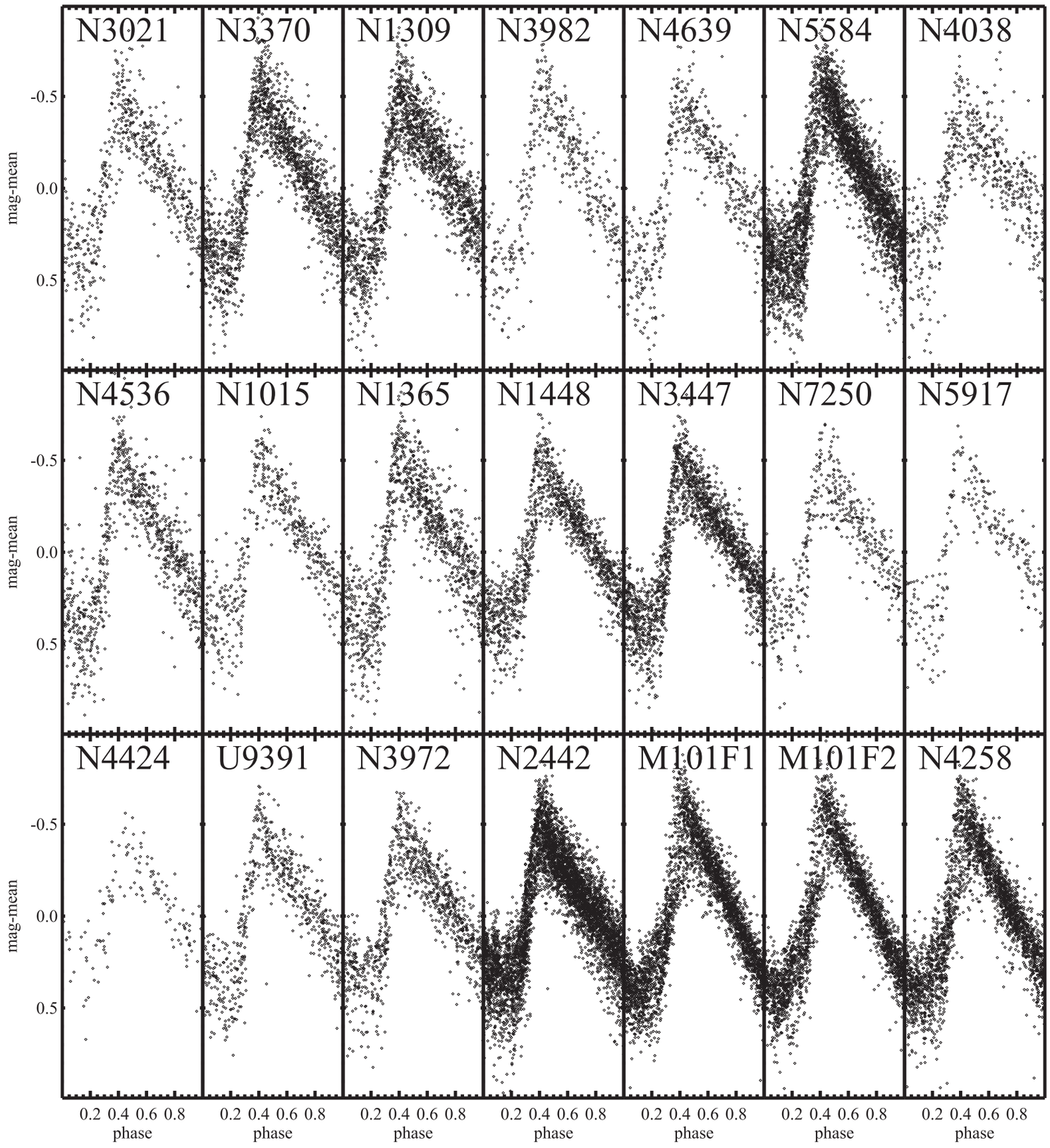
- Fitting Statistical Models to Astronomical Data
    - Linear Regression Approaches (F&B Ch 7, Ivezic, Ch 8)
    - Generative / Forward Modeling with Latent Variables
    - Linear Regression with intrinsic dispersion and heteroskedastic (x,y) measurement error
        - Kelly et al. “Some Aspects of Measurement Error in Linear Regression of Astronomical Data.” 2017, The Astrophysical Journal, 665, 1489
- Bayesian Inference in Astronomy (F&B 3.8, Ivezic 5)
    - C. Bailer-Jones.“Estimating Distances from Parallaxes.” 2015, PASP,127, 994 https://arxiv.org/abs/1507.02105

- Looking ahead:
  - Review generation of random numbers for arbitrary probability distributions (Ivezic 3.7)
  - Patel, Besla & Mandel. “Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations.” MNRAS, 468, 3428. https://arxiv.org/abs/1803.01878
  - Patel, Besla, Mandel & Sohn. "Estimating the Mass of the Milky Way Using the Ensemble of Classical Satellite Galaxies.” The Astrophysical Journal, 857, 78. https://arxiv.org/abs/1703.05767

Regression

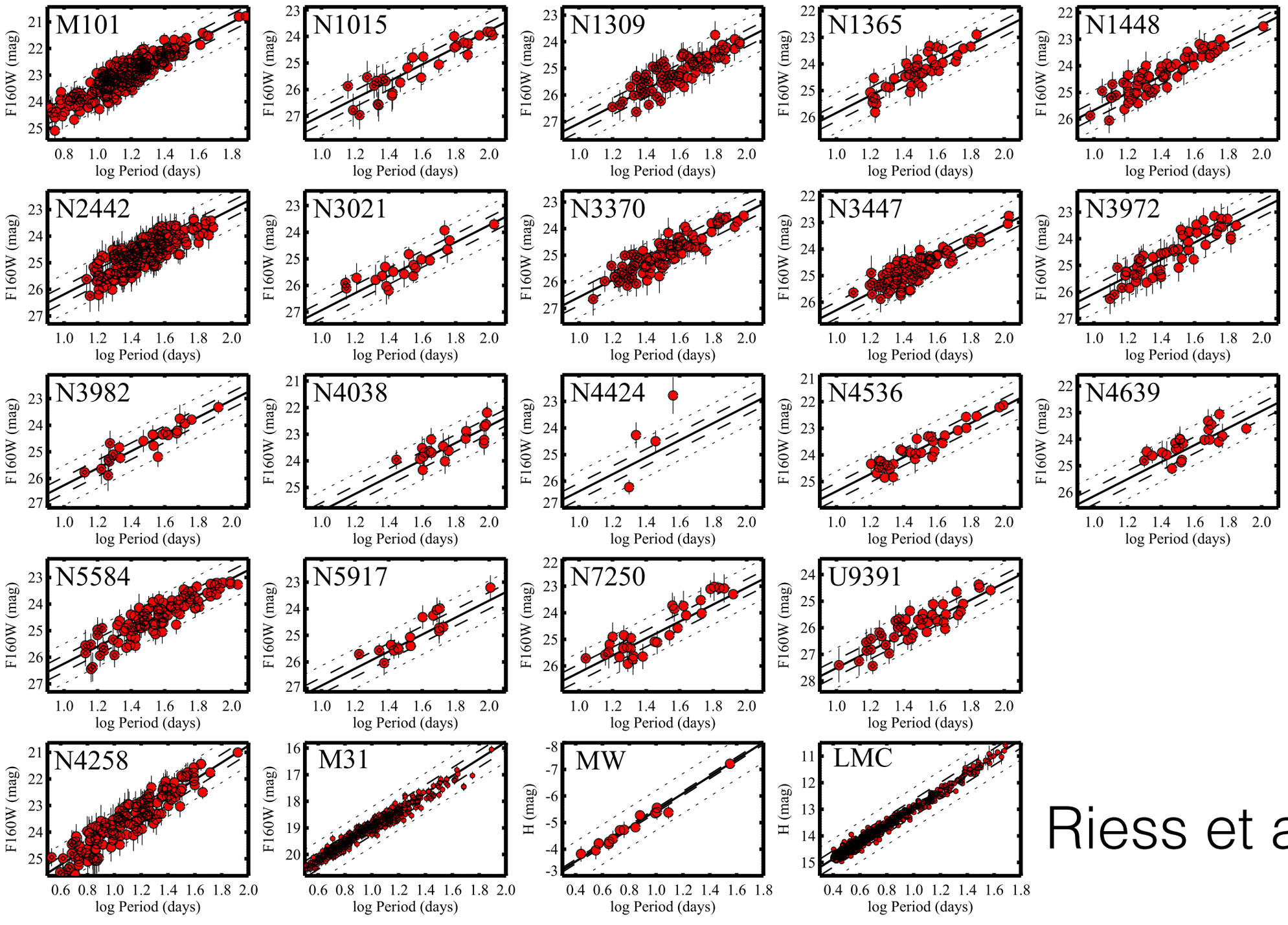
- Fitting a function E[y | x] = f(x; θ) for the mean relation between y and x
- Zoo of Methods (see F&B Ch 7, Ivezic Ch 8)
- Basic Approaches
  - Ordinary Least Squares (homoskedastic scatter)
  - Generalised Least Squares (heteroskedastic, correlated scatter)
  - Weighted Least Squares (minimum χ^2) (known variance)
  - Maximum Likelihood
- Real data problems require more complex statistical modelling

Example: Cepheid Light Curves (Time Series)



Riess et al. 2016

Example: Leavitt's Law: Period-Luminosity Relation



Riess et al. 2016

Figure 6. Near-infrared Cepheid *P–L* relations. The Cepheid magnitudes are shown for the 19 SN hosts and the four distance-scale anchors. Magnitudes labeled as *F160W* are all from the same instrument and camera, WFC3 *F160W*. The uniformity of the photometry and metallicity reduces systematic errors along the distance ladder. A single slope is shown to illustrate the relations, but we also allow for a break (two slopes) as well as limited period ranges.

Ordinary Least Squares (OLS)

Linear Model

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \, x_{ij} + \epsilon_i$$

$$i = 1, \ldots, N \text{ objects}$$

$$\mathbb{E}[\epsilon_i] = 0$$

homoskedastic

$$\mathrm{Var}[\epsilon_i] = \sigma^2 \text{ (known)}$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Minimise wrt β:

$$\mathrm{RSS} = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j}^{k-1} \beta_j \, x_{ij})^2 = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\hat{\beta}_{\mathrm{OLS}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y} \qquad \mathrm{Var}[\hat{\beta}_{\mathrm{OLS}}] = \sigma^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1}$$

$$\mathbb{E}[\hat{\beta}_{\mathrm{OLS}}] = \beta(\text{ unbiased}) \qquad \text{BLUE}$$

Ordinary Least Squares (OLS)

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \, x_{ij} + \epsilon_i$$

$$i = 1, \ldots, N \text{ objects}$$

$$\mathbb{E}[\epsilon_i] = 0$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathrm{Var}[\epsilon_i] = \sigma^2(\text{ unknown})$$

Estimate unknown variance as:

$$\widehat{\sigma^2} = \frac{1}{N-k}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})^T(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})$$

Weighted Least Squares aka χ^2 minimisation

Linear Model

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \, x_{ij} + \epsilon_i$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$i = 1, \ldots, N \text{ objects}$$

$$\mathbb{E}[\epsilon_i] = 0$$

heteroskedastic

$$\mathrm{Var}(\epsilon_i) = \sigma_i^2 (\text{ known})$$

Minimise wrt β:

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - \beta_0 - \sum_j^{k-1} \beta_j \, x_{ij})^2}{\sigma_i^2}$$

χ^2 r.v. = sum of squared Gaussian r.v.s

If Gaussian errors, at $\beta = \beta_{min}$ $\quad X^2 \sim \chi^2_{N-k}$

model check: $\mathbb{E}(\chi^2_{N-k}) = N - k$ $\quad \dfrac{X^2}{N-k} \approx 1 (\text{ for large N-k})$

These are special cases of Generalised Least Squares

Linear Model

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \, x_{ij} + \epsilon_i$$

$$i = 1, \ldots, N \text{ objects}$$

$$\mathbb{E}[\epsilon_i] = 0$$

Correlated Errors

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathrm{Var}[\boldsymbol{\epsilon}] = \mathrm{Cov}[\boldsymbol{\epsilon}, \boldsymbol{\epsilon}^T] = \boldsymbol{W} (\text{known})$$

Minimise wrt β:

$$\mathrm{RSS} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{W}^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\widehat{\beta}_{\mathrm{GLS}} = (\boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}^{-1} \boldsymbol{Y}$$

$$\mathbb{E}[\widehat{\beta}_{\mathrm{GLS}}] = \beta (\text{unbiased}) \qquad \mathrm{Var}[\widehat{\beta}_{\mathrm{GLS}}] = (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1}$$

They can also be thought of as Maximum Likelihood (assuming Gaussian errors)

Linear Model

$$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_j \, x_{ij} + \epsilon_i$$

$$i = 1, \ldots, N \text{ objects}$$

Correlated Errors

$$\boldsymbol{\epsilon} \sim N(0, \boldsymbol{W})$$

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{W})$$

Maximise wrt β:

$$L(\boldsymbol{\beta}) = P(\boldsymbol{Y}|\,\boldsymbol{\beta}, \boldsymbol{X}) = N(\boldsymbol{Y}|\,\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{W})$$

Fitting Models to Astro Data

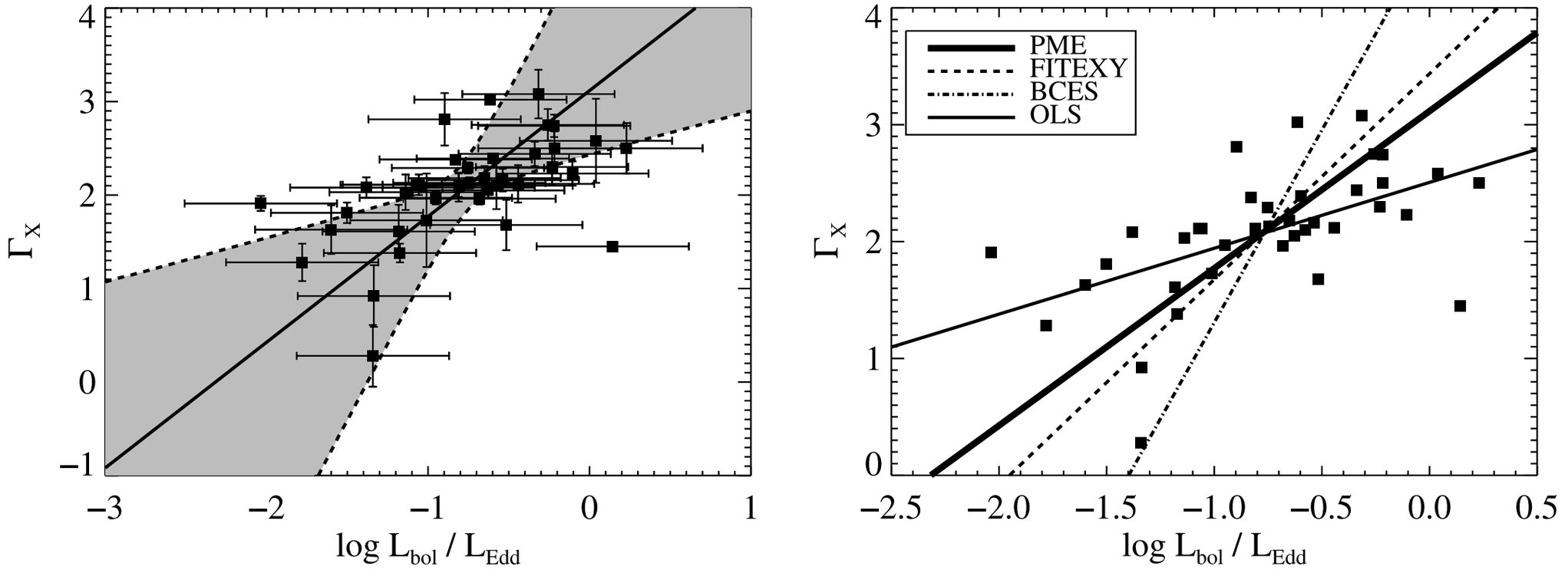


FIG. 10.—X-ray photon index Γ_X as a function of $\log L_{bol}/L_{Edd}$ for 39 $z \lesssim 0.8$ radio-quiet quasars. In both plots, the thick solid line shows the posterior median estimate (PME) of the regression line. In the left panel, the shaded region denotes the 95% (2 σ) pointwise confidence intervals on the regression line. In the right panel, the thin solid line shows the OLS estimate, the dashed line shows the FITEXY estimate, and the dot-dashed line shows the BCES($Y|X$) estimate; the error bars have been omitted for clarity. A significant positive trend is implied by the data.

Modelling heteroskedastic, correlated measurement errors in both y and x, intrinsic scatter, nondetections, selection effects

B. Kelly et al. 2007, “Some Aspects of Measurement Error in Linear Regression of Astronomical Data.” ApJ, 665, 1489

Ad-hoc "χ^2" approaches vs. Likelihood formulation

FITEXY Estimator

- Press et al.(1992, *Numerical Recipes*) define an 'effective χ^2' statistic:

$$\chi^2_{EXY} = \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2_{y,i} + \beta^2 \sigma^2_{x,i}}$$

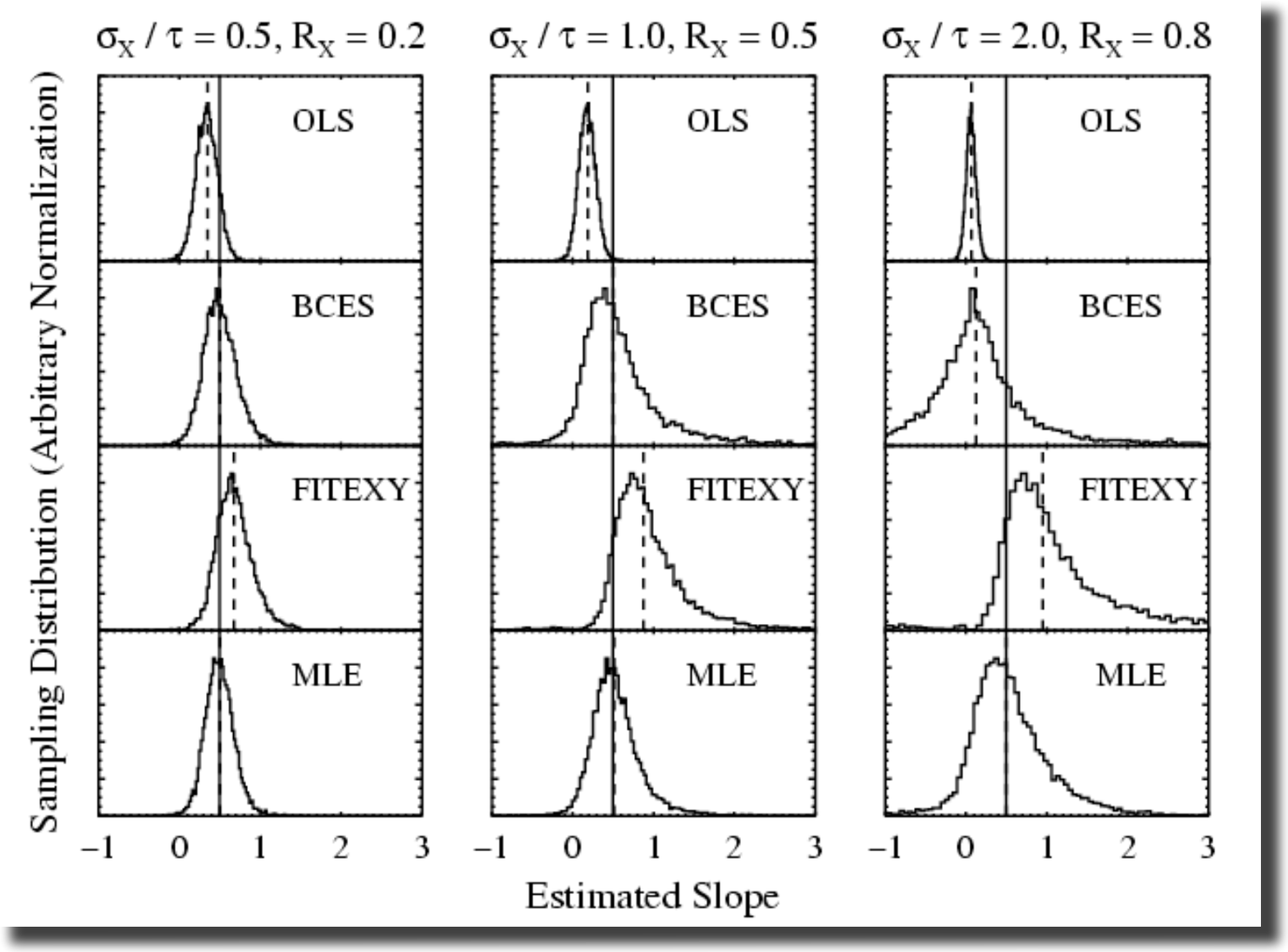
- Choose values of α and β that minimize χ^2_{EXY}
- Modified by Tremaine et al.(2002, ApJ, 574, 740), to account for intrinsic scatter: **σ**

$$\chi^2_{EXY} = \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma^2 + \sigma^2_{y,i} + \beta^2 \sigma^2_{x,i}}$$

http://astrostatistics.psu.edu/su07/kelley_measerr07.pdf

Kelly et al. 2017, Latent Variable Likelihood approach vs. Bad Approaches

Simulation Study: Slope



Dashed lines mark the median value of the estimator, solid lines mark the true value of the slope. Each simulated data set had 50 data points, and y-measurement errors of $\sigma_y \sim \sigma$.

http://astrostatistics.psu.edu/su07/kelley_measerr07.pdf

Probabilistic Generative Modelling

- Forward Model comprises series of probabilistic steps describing conceptually how the observed data was generated from the parameters of interest

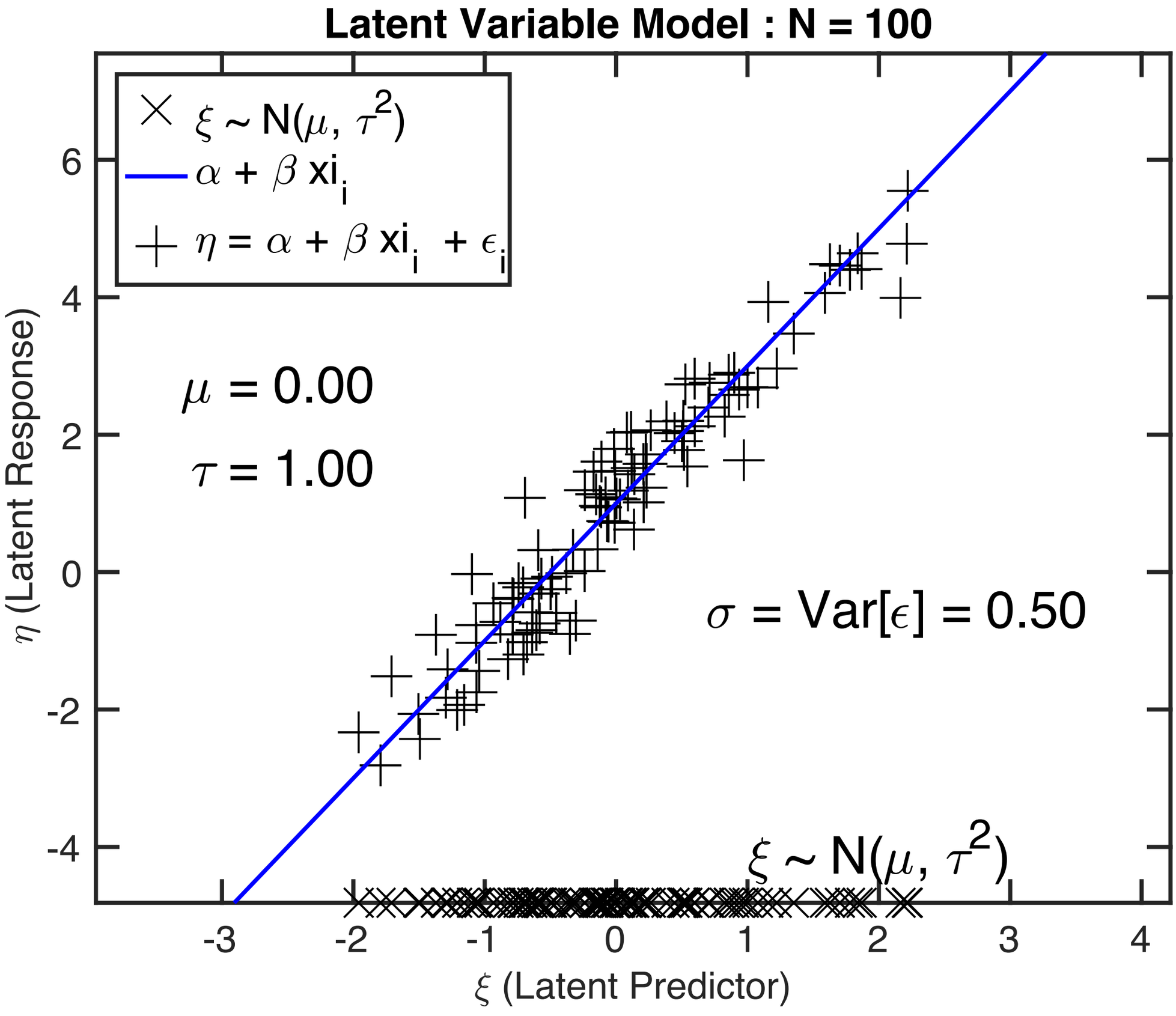
- Can introduce intermediate parameters / unobserved latent variables $\boldsymbol{\alpha}$ (e.g. true values corresponding to the observed data).

- From Forward model, derive the sampling distribution, e.g.
  $P(D \mid \theta) = \int P(D \mid \boldsymbol{\alpha})\, P(\boldsymbol{\alpha} \mid \theta)\, d\boldsymbol{\alpha}$

- Using observed data D, draw inference from Likelihood function:
  $L(\theta) = P(D \mid \theta)$

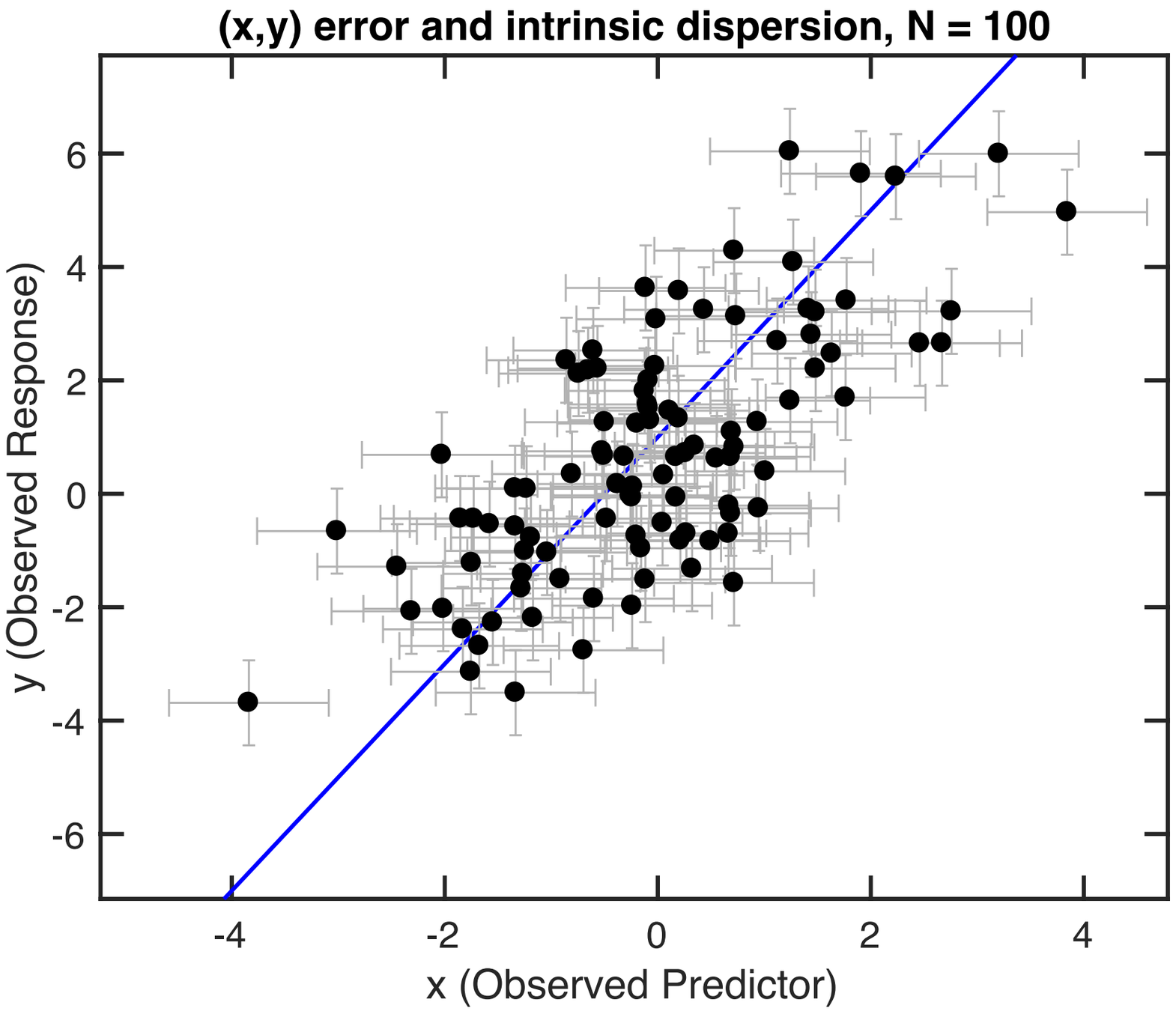
- Or if Bayesian with prior $P(\theta)$: sample posterior:
  $P(\theta \mid D) = P(D \mid \theta)\, P(\theta)$

Latent Variable Model



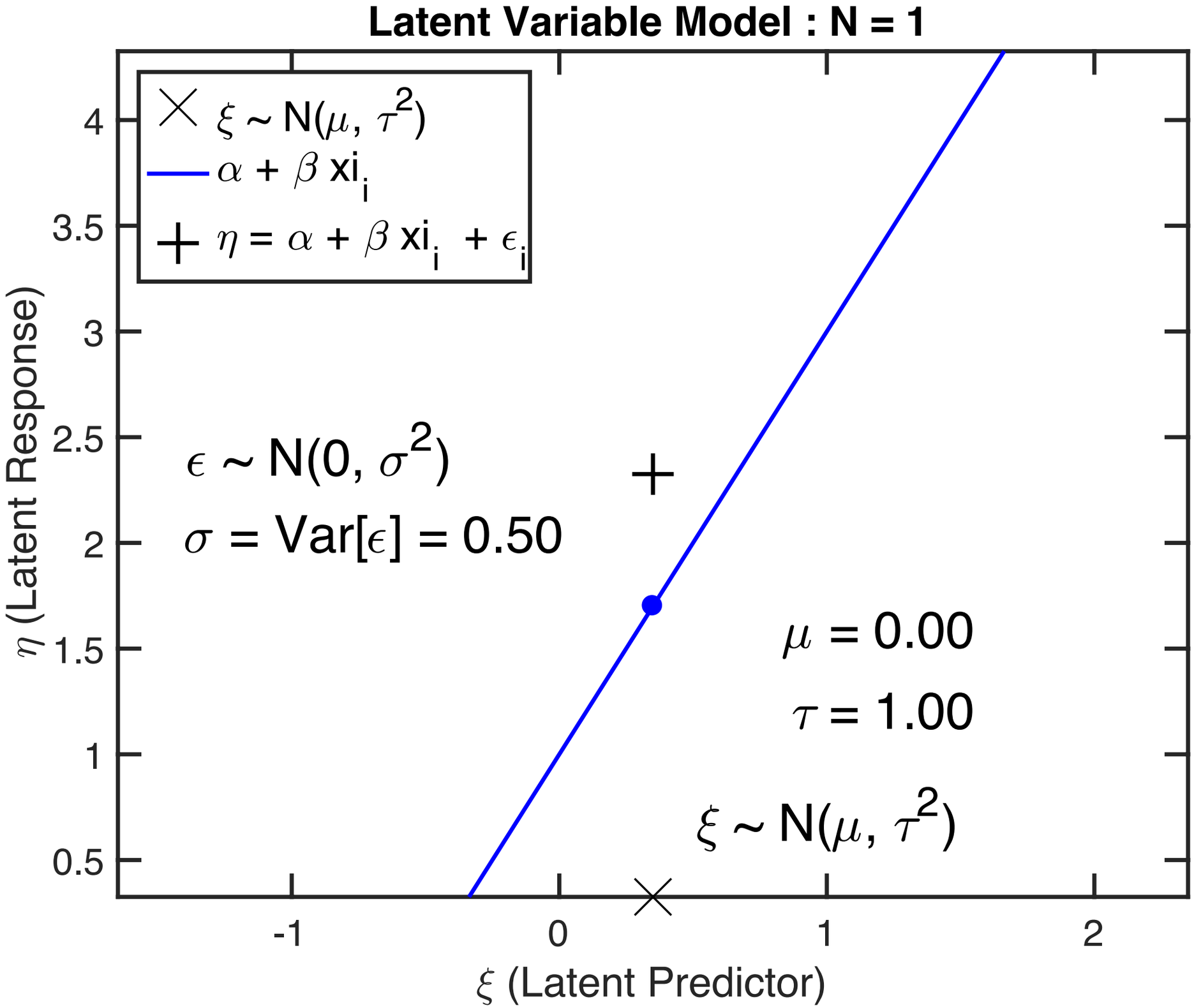
Example: Structural Model for Linear Regression (B. Kelly et al. 2007, “Some Aspects of Measurement Error in Linear Regression of Astronomical Data.” ApJ, 665, 1489)

- Observed data has x and y meas. errors and intrinsic dispersion
- Estimate the true slope (and other parameters)



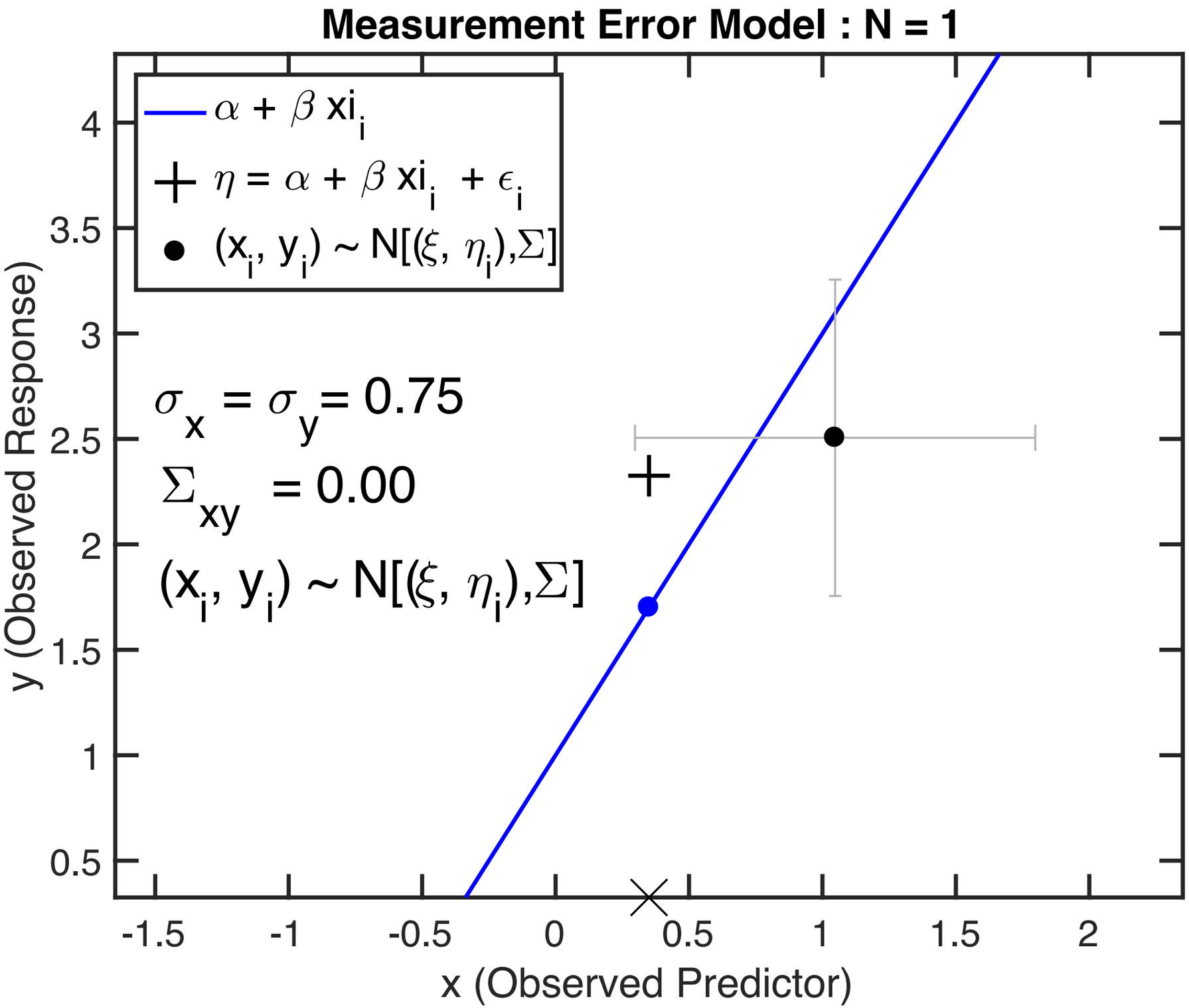
Step 1: Generating Latent Variables from Parameters:

$$P(\eta_i, \xi_i | \alpha, \beta, \sigma, \mu, \tau) = P(\eta_i | \xi_i, \alpha, \beta, \sigma) \times P(\xi_i | \mu, \tau)$$

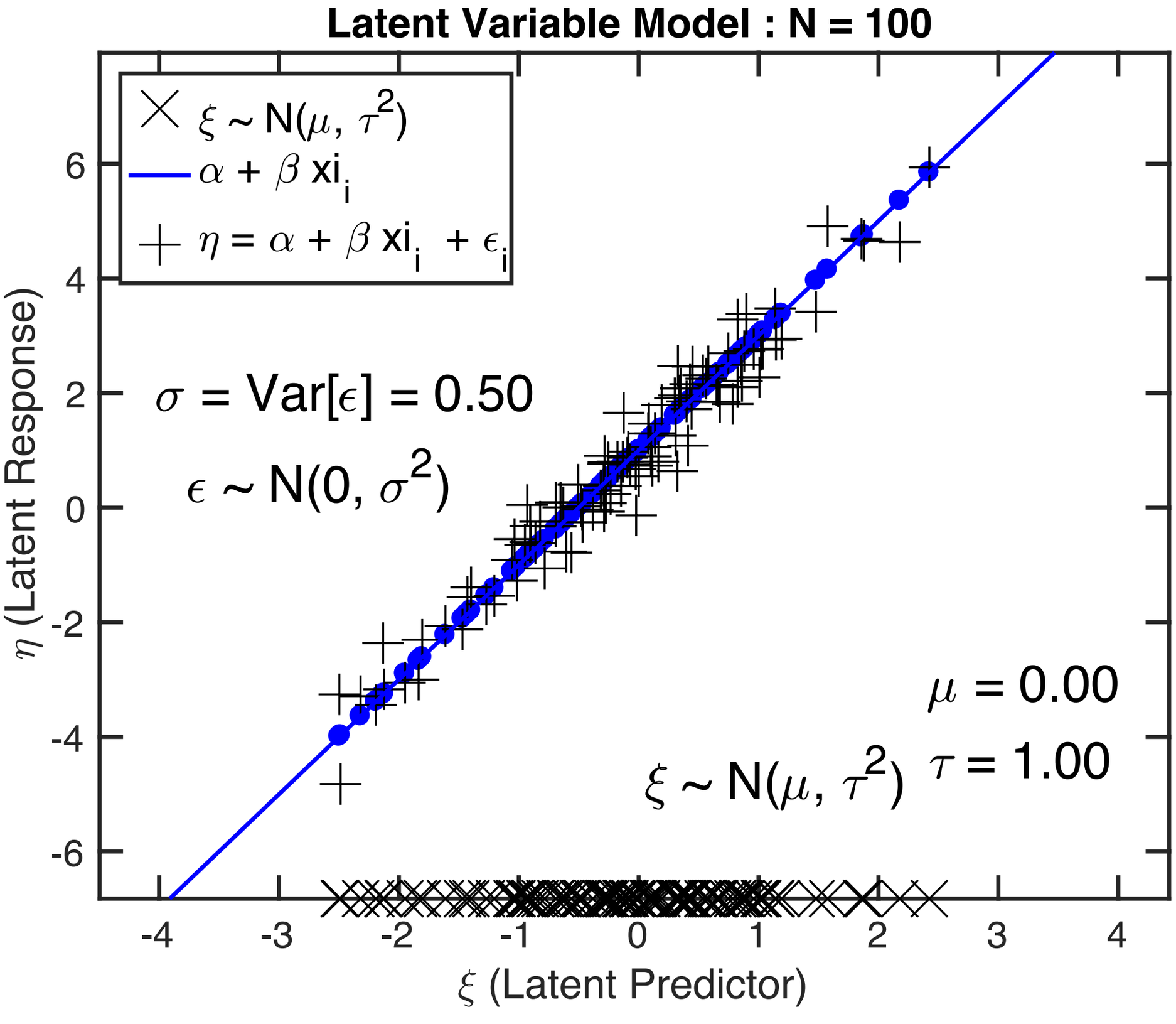


Step 2: Generating Observed Data from Latent Variables

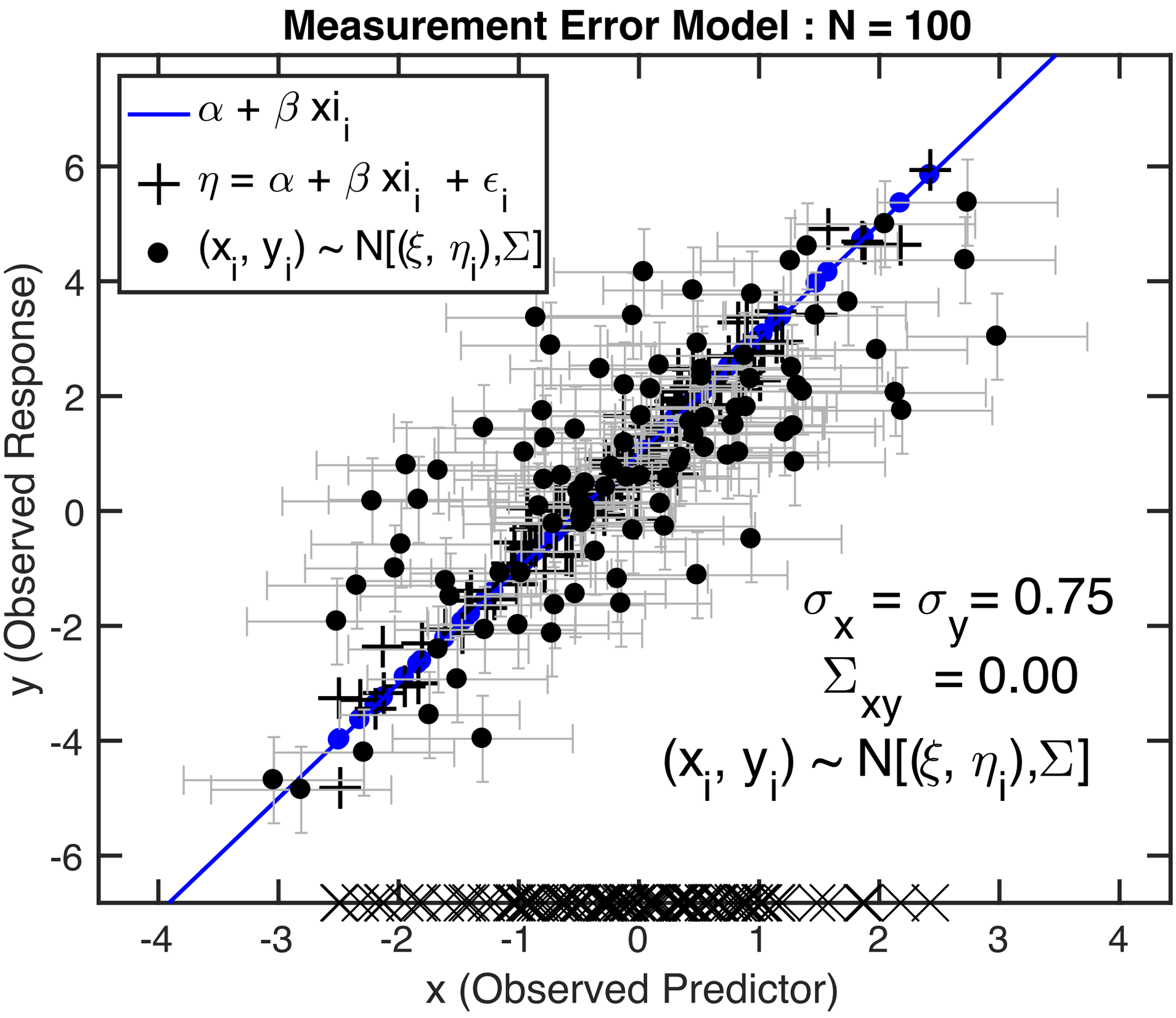
$$P([x_i, y_i]|\eta_i, \xi_i) = N([x_i, y_i]|[\eta_i, \xi_i], \boldsymbol{\Sigma})$$



Now repeat for N=100 objects



Now repeat for N=100 objects



Knowns and Unknowns

Regression Parameters $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$

Independent Variable
Population Distribution
"Hyperparameters" $\boldsymbol{\psi} = (\mu, \tau)$

Latent (true) Variables (ξ_i, η_i)

Observed Data (x_i, y_i)

Generative Model

Population Distribution

$$\xi \sim N(\mu|\tau^2)$$

Regression

$$\eta_i|\,\xi_i \sim N(\alpha + \beta\xi_i, \sigma^2)$$

Measurement Error

$$[x_i, y_i]|\xi_i, \eta_i \sim N([\xi_i, \eta_i], \Sigma])$$

Formulating Likelihood Function: Marginalising (integrating out) latent variables

$$P(x_i, y_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \int\int P(x_i, y_i, \xi_i, \eta_i|\boldsymbol{\theta}, \boldsymbol{\psi}) d\xi_i \, d\eta,$$

$$P(x_i, y_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \int\int P(x_i, y_i|\xi_i, \eta_i) P(\eta_i|\xi_i, \boldsymbol{\theta}) P(\xi_i|\boldsymbol{\psi}) \, d\xi_i \, d\eta$$

Measurement Error

Regression

Population Distribution of Covariate

Example Sheet 2: Derive this!

Solution: (Kelly 2007, Eqs. 16-23)

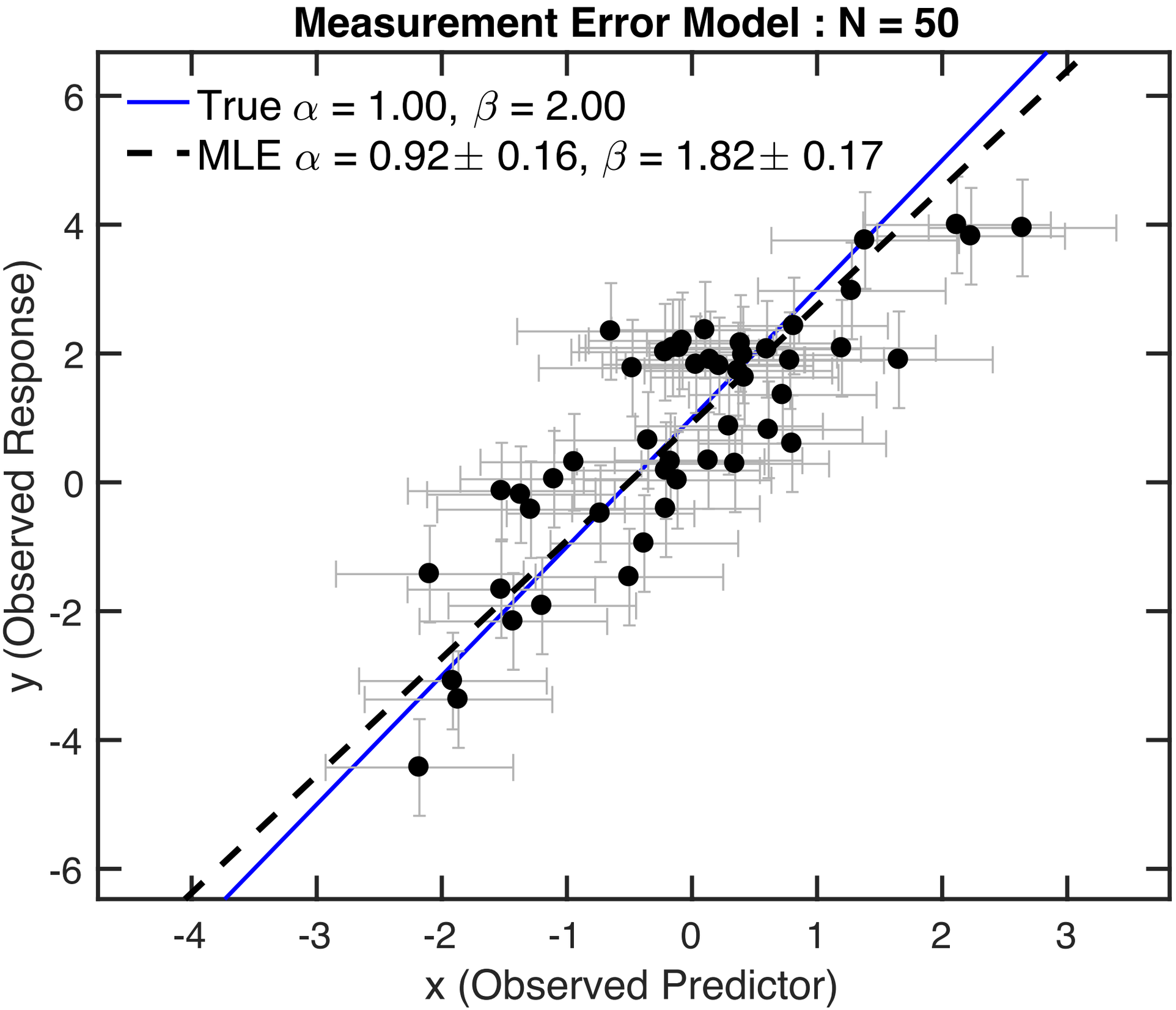
Gave More General Solution when P(ξ|Ψ)
is a Mixture of Gaussians
(set K=1, $\pi_1 = 1$ for us)

$$p(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\psi}) = \prod_{i=1}^{n}\sum_{k=1}^{K}\frac{\pi_k}{2\pi|\mathbf{V}_{k,i}|^{1/2}} \times \exp\left[-\frac{1}{2}(\boldsymbol{z}_i-\boldsymbol{\zeta}_k)^T\mathbf{V}_{k,i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\zeta}_k)\right], \quad (16)$$

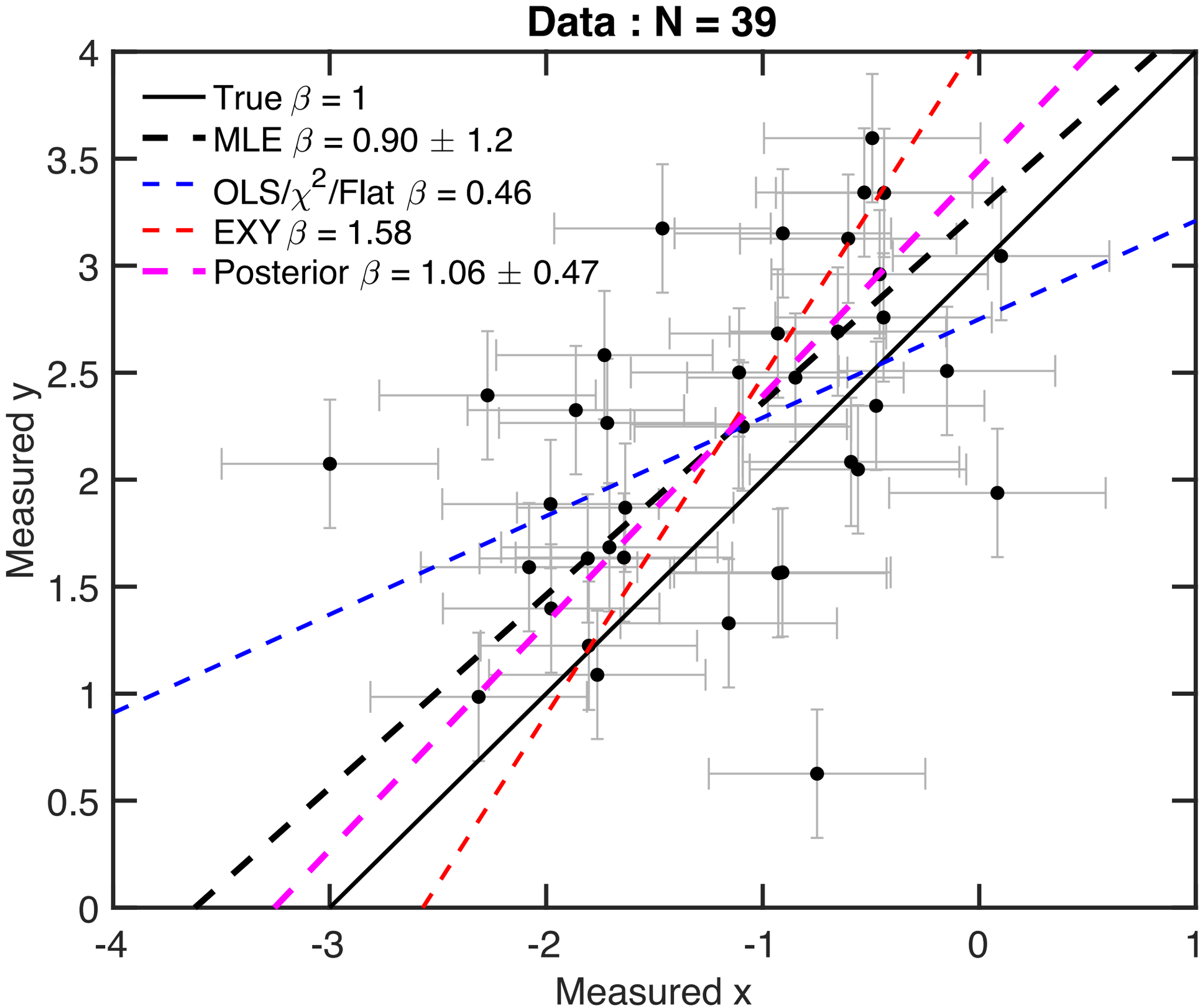
$$\boldsymbol{\zeta}_k = (\alpha+\beta\mu_k,\mu_k), \quad (17)$$

$$\mathbf{V}_{k,i} = \begin{pmatrix}\beta^2\tau_k^2+\sigma^2+\sigma_{y,i}^2 & \beta\tau_k^2+\sigma_{xy,i}\\ \beta\tau_k^2+\sigma_{xy,i} & \tau_k^2+\sigma_{x,i}^2\end{pmatrix}, \quad (18)$$

Example

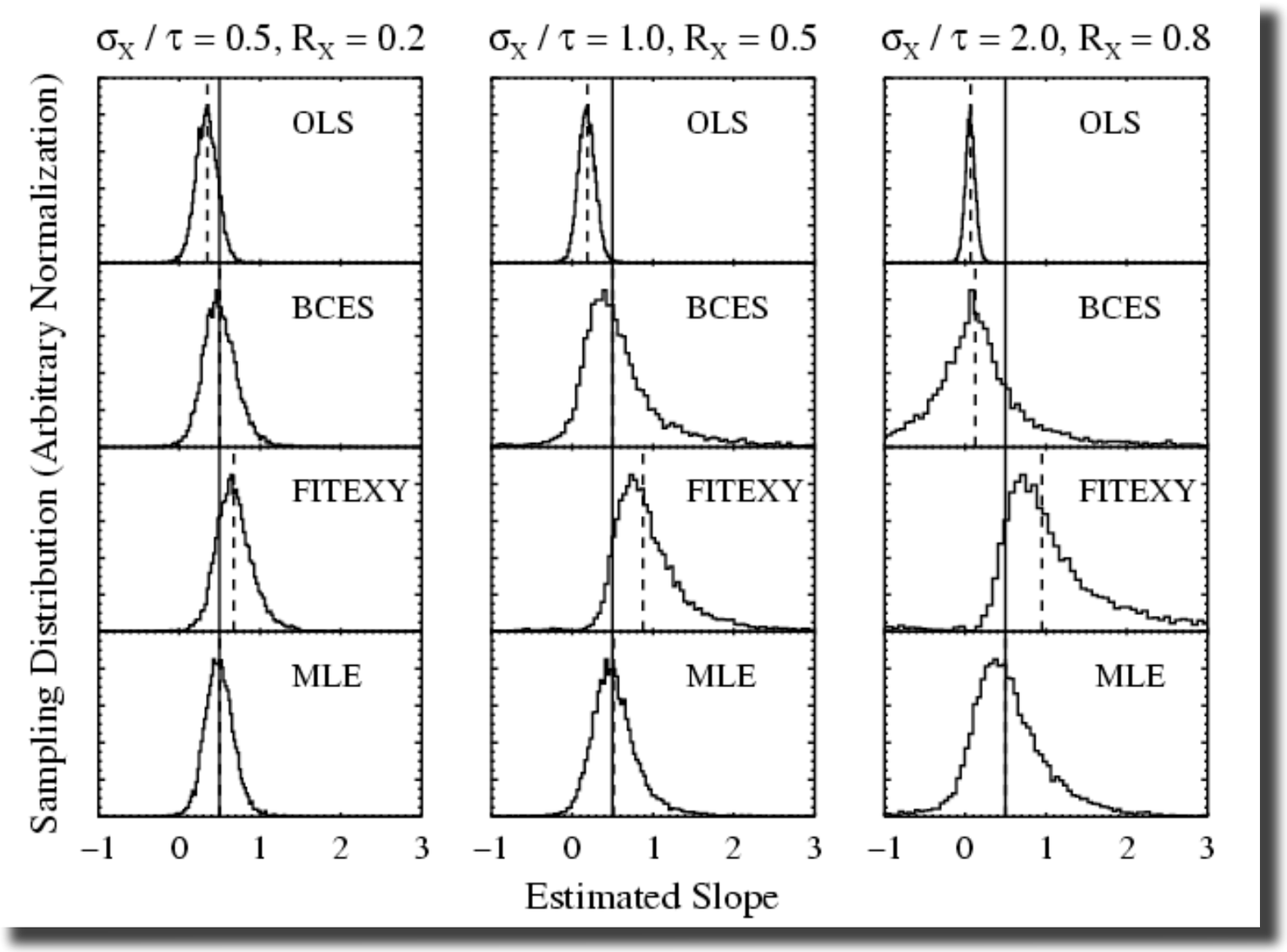

Measurement Error Model : N = 50
True α = 1.00, β = 2.00
MLE α = 0.92± 0.16, β = 1.82± 0.17
y (Observed Response)
x (Observed Predictor)
6
4
2
0
-2
-4
-6
-4
-3
-2
-1
0
1
2
3

Example Sheet: OLS is biased towards lower slopes, does not account for x-measurement error



Kelly et al. 2017, Latent Variable Likelihood approach vs. Bad

Simulation Study: Slope



Dashed lines mark the median value of the estimator, solid lines mark the true value of the slope. Each simulated data set had 50 data points, and y-measurement errors of $\sigma_y \sim \sigma$.

http://astrostatistics.psu.edu/su07/kelley_measerr07.pdf

Statistical Modelling Wisdom

- Have an objective function [e.g. Likelihood or posterior] that you optimise or sample to fit the data - not just a procedure/recipe

- Objective function helps you evaluate relative fits of data with under different parameter values / models

- Derive your objective function from your modelling assumptions (physical or statistical)

- Write down your assumptions!

- First question: what is the likelihood L(θ) ? Derive it from the assumptions underlying your sampling distribution P(D | θ)!

- Second question: what is your prior P(θ) ? (if Bayesian)

- Third question: How do I optimise/sample objective function to fit the data?