Astrostatistics: Fri 08 Feb 2019

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

- 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

Astrostatistics: Fri 08 Feb 2019

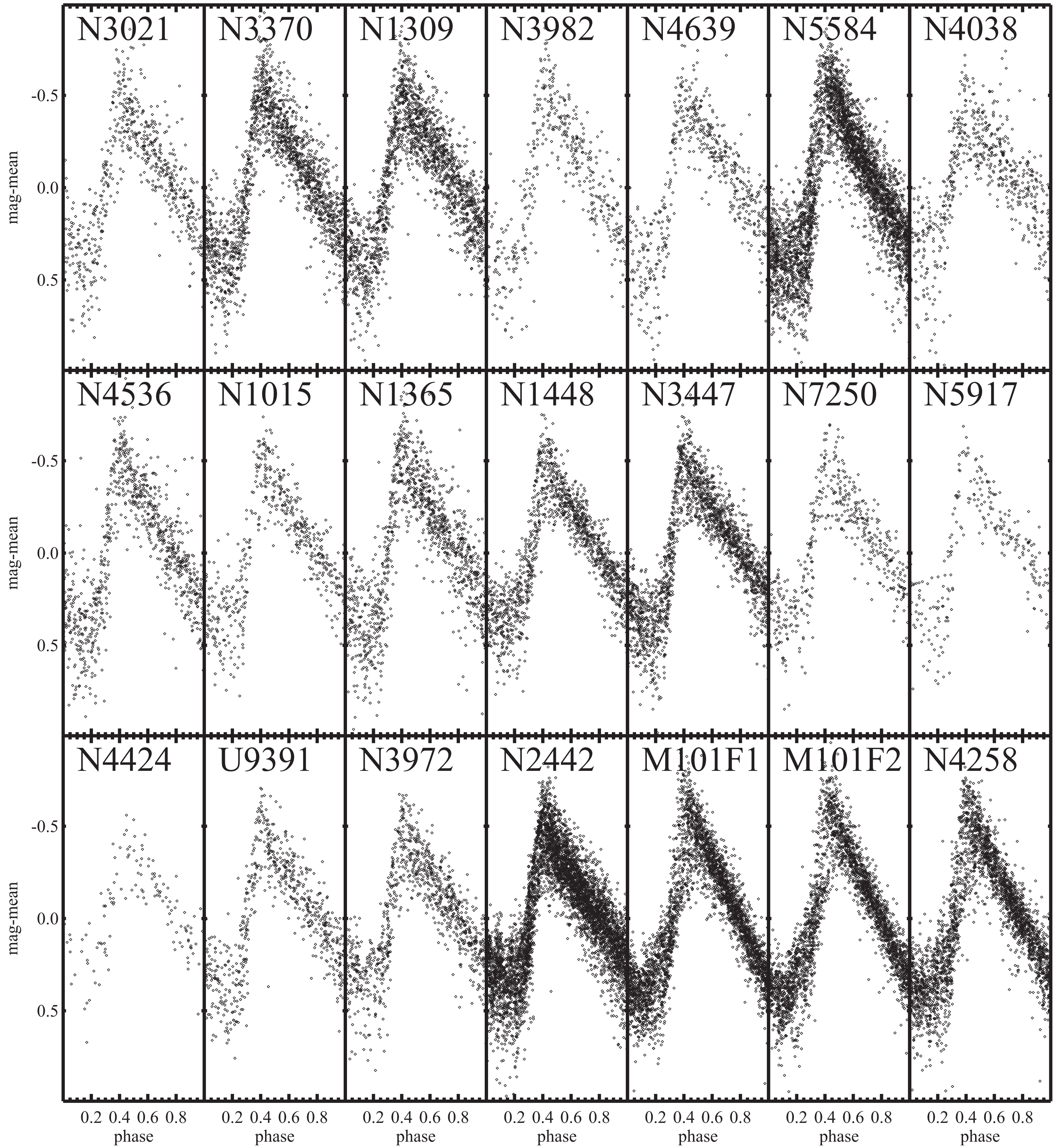
https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

- 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

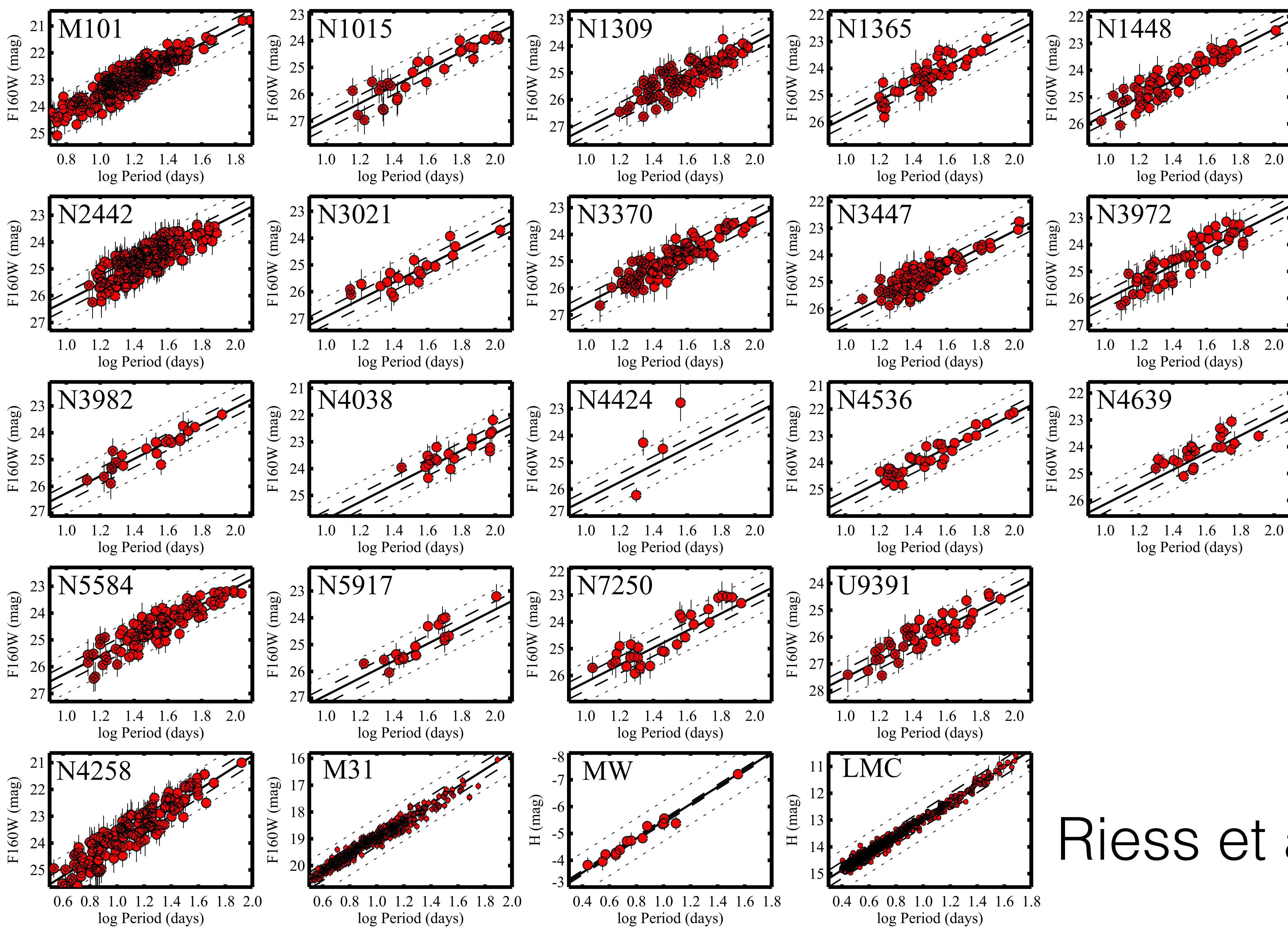
- Fitting a function E[y | x] = f(x; θ) for the mean relation between y and x
- Zoo of Methods (see F&B Ch 7, Ivezic Ch 8)
- Review Basic Approaches
    - Ordinary Least Squares (homoskedastic scatter)
    - Generalised Least Squares (heteroskedastic, correlated scatter)
    - Weighted Least Squares (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.

Review: Ordinary Least Squares

Linear Model

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

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

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

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

homoskedastic

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

Minimise wrt β:

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

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

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

BLUE

Review: Ordinary Least Squares

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

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

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

$$\downarrow$$

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

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

Estimate unknown variance as:

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

Review: Weighted Least Squares aka χ2 minimisation

Linear Model

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

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

$i = 1, \ldots, N$ 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 $\beta = \beta_{min}$

$$X^2 \sim \chi^2_{N-k}$$

model check: $\mathbb{E}(\chi^2_{N-k}) = N - k$

$$\frac{X^2}{N-k} \approx 1 (\text{ for large N-k})$$

These are special cases of Generalised Least Squares

Linear Model

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

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

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

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

Correlated Errors

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

<u>Minimise wrt β:</u>

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

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

$$\mathbb{E}[\widehat{\beta}_{\mathrm{GLS}}] = \beta (\text{unbiased})$$

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

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

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

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

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

<u>Maximise wrt β:</u>

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

Fitting Models to Astro Data

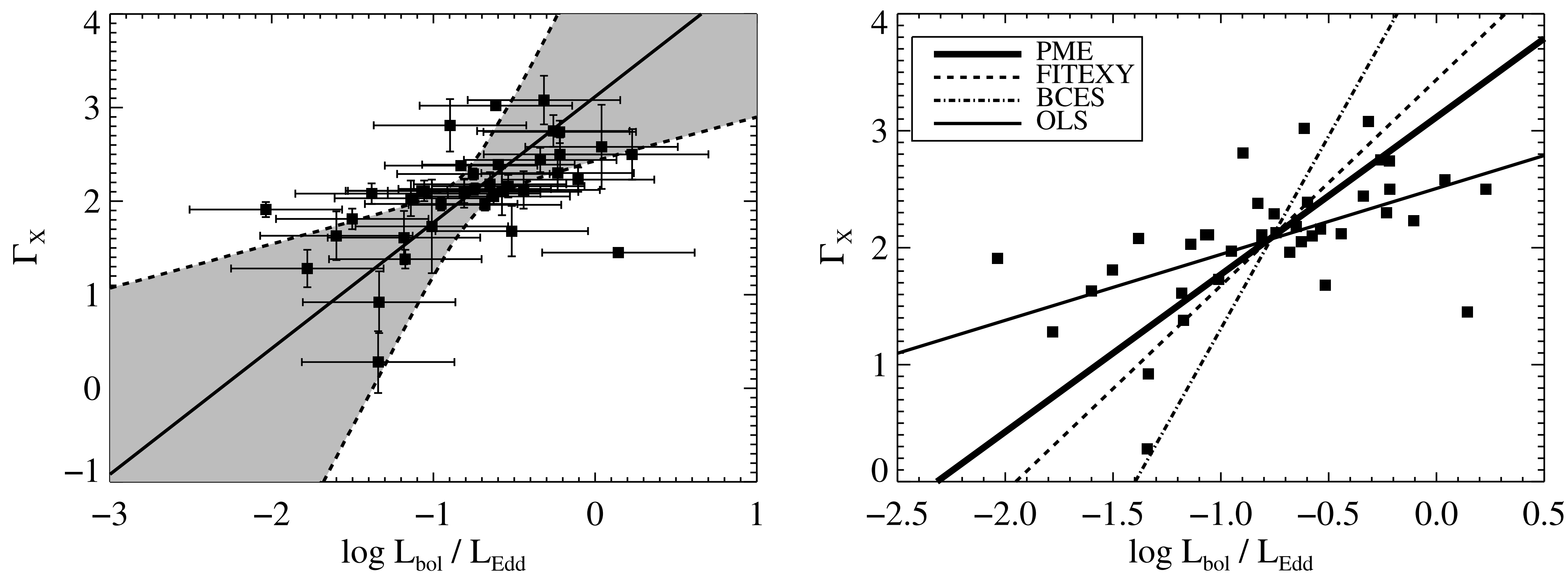


FIG. 10.—X-ray photon index Γ_X as a function of $\log L_{bol}/L_{Edd}$ for 39 $z \lesssim 0.8$ radio-quiet quasars. In both plots, the thick solid line shows the posterior median estimate (PME) of the regression line. In the left panel, the shaded region denotes the 95% ($2\,\sigma$) pointwise confidence 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:

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

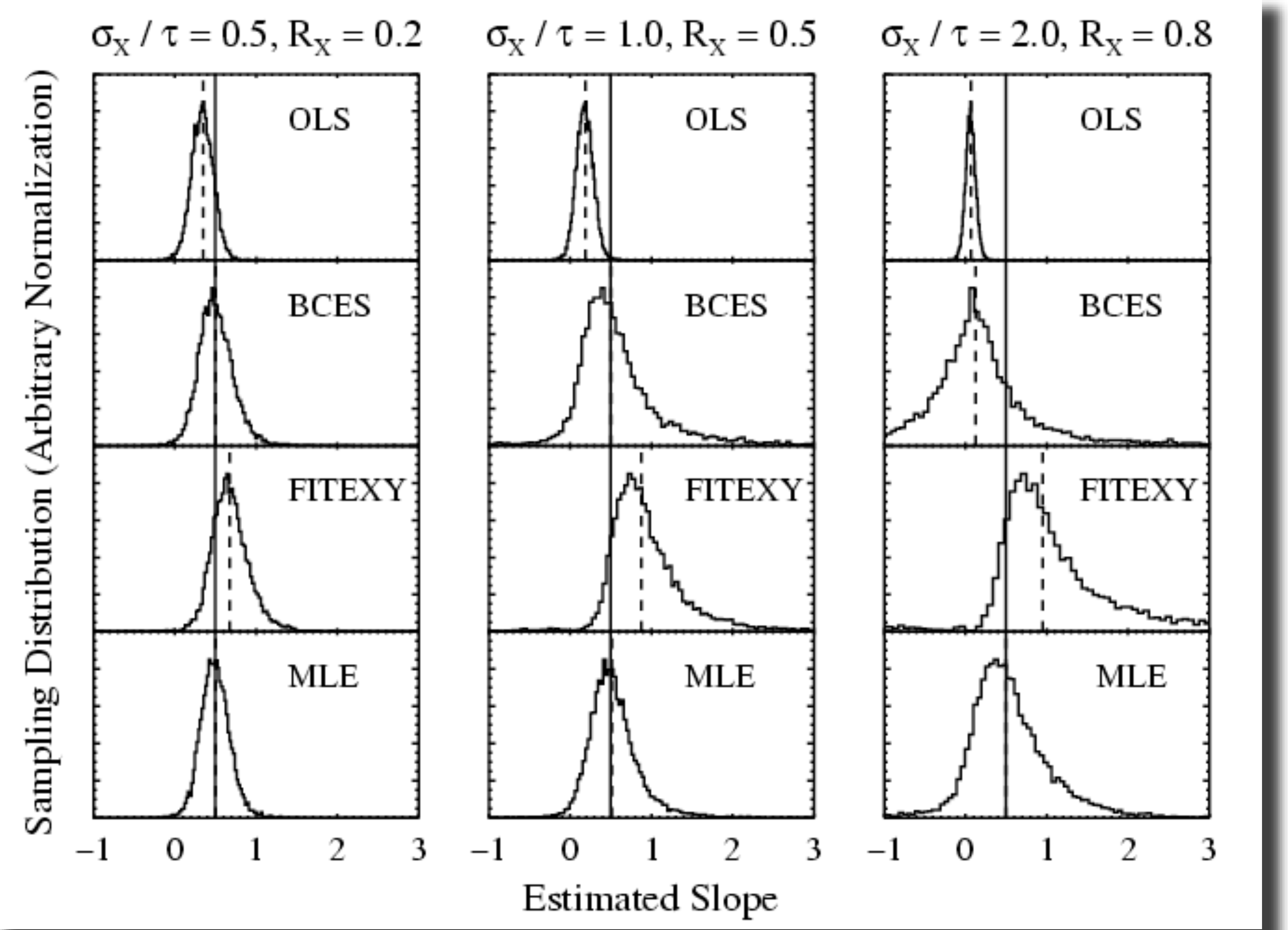
- Choose values of α and β that minimize χ^2_{EXY}
- Modified by Tremaine et al.(2002, ApJ, 574, 740), to account for intrinsic scatter:

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

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

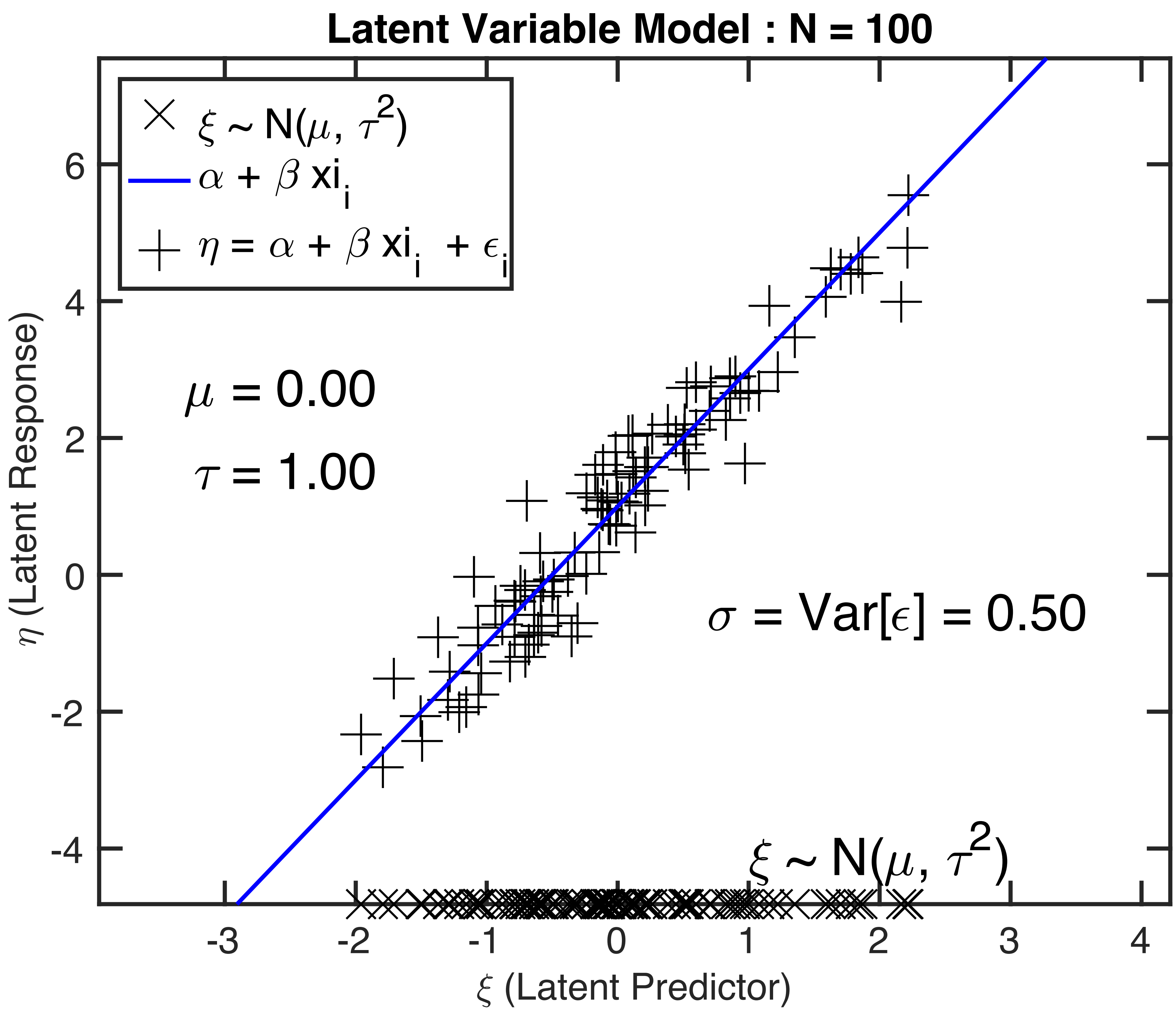
- Can introduce intermediate parameters / unobserved latent variables $\boldsymbol{\alpha}$ (e.g. true values corresponding to the observed data).

- 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:
  $P(\theta \mid D) = P(D \mid \theta)\, P(\theta)$

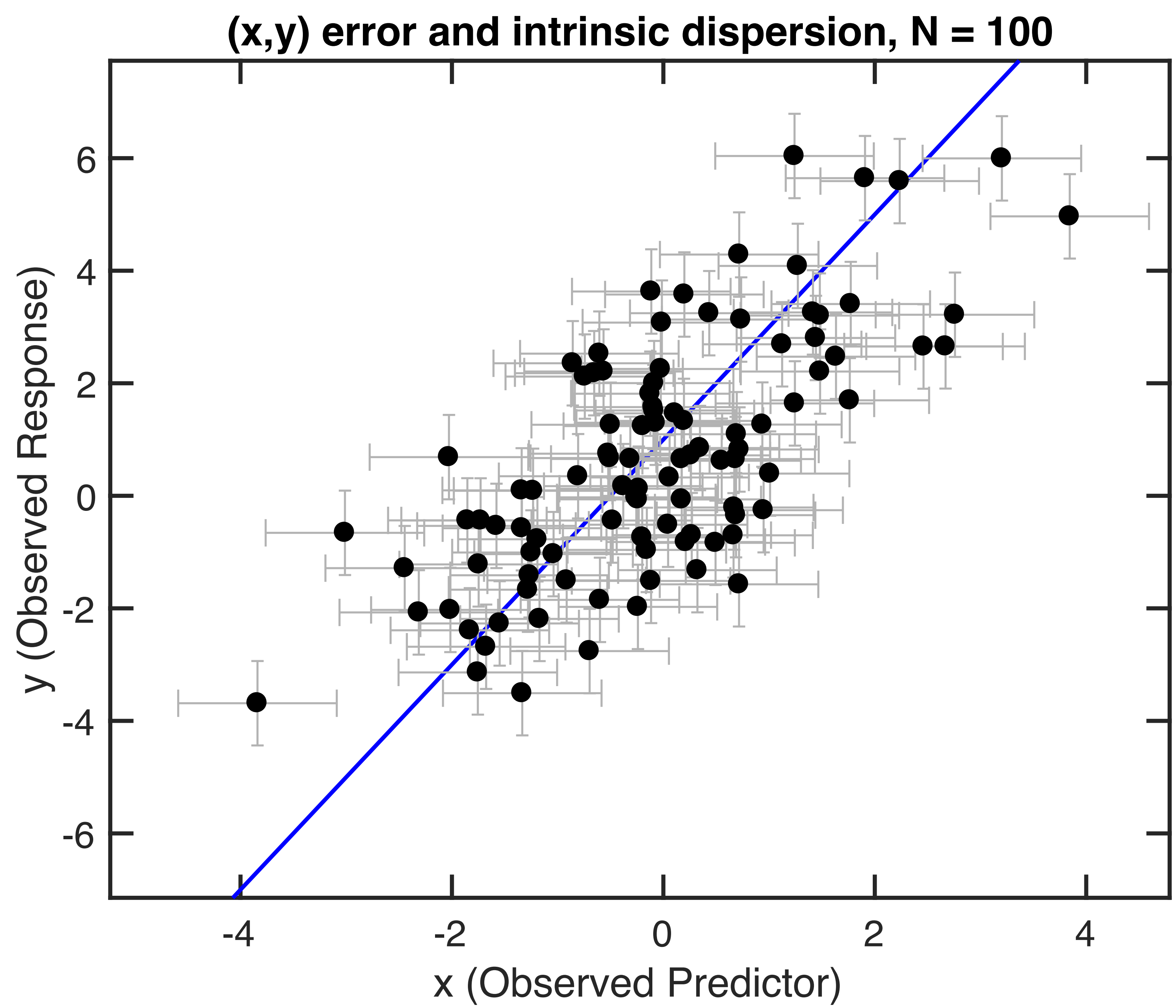
Latent Variable Model



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)

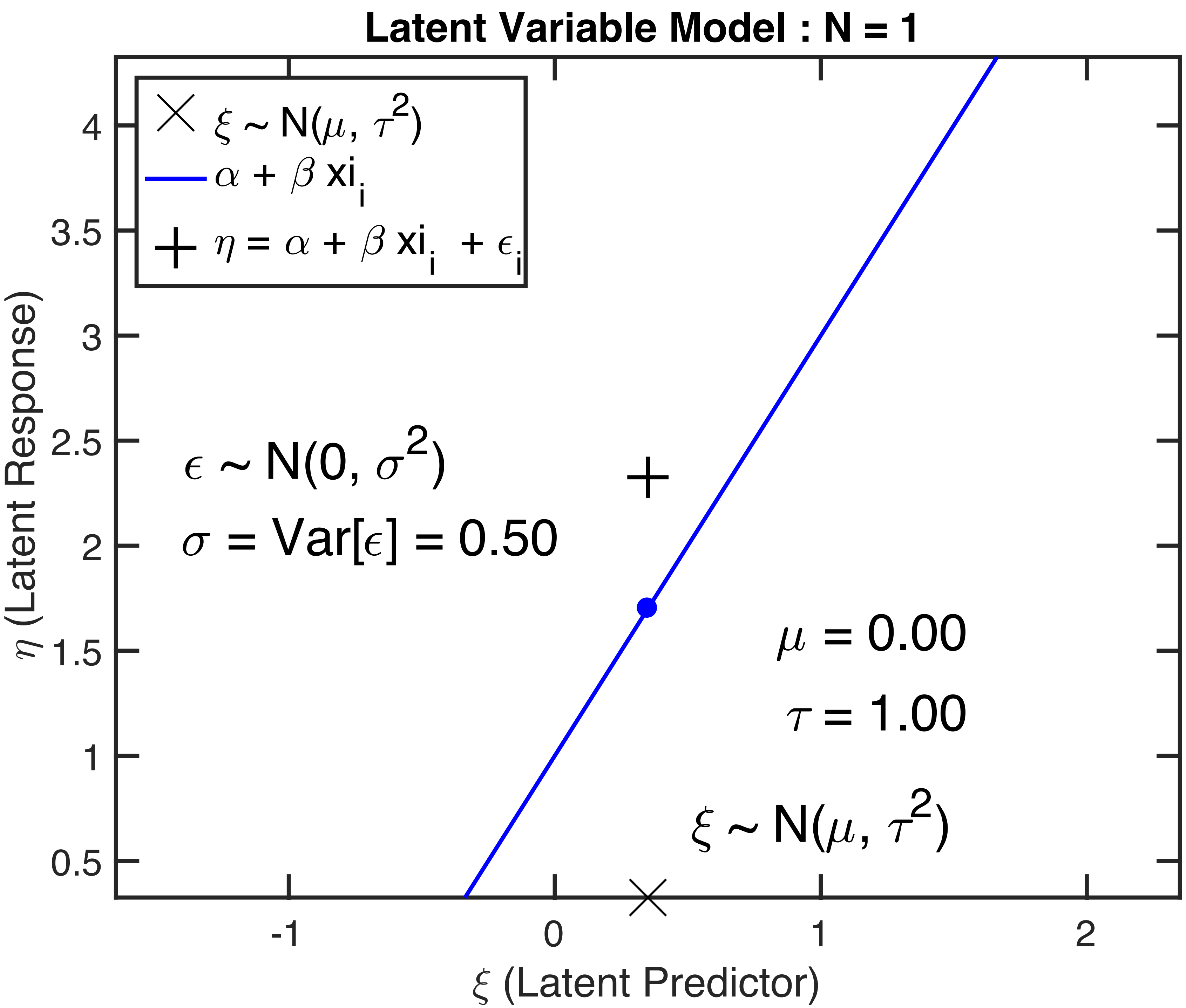
- Observed data has x and y meas. errors and intrinsic dispersion

- Estimate the true slope (and other parameters)



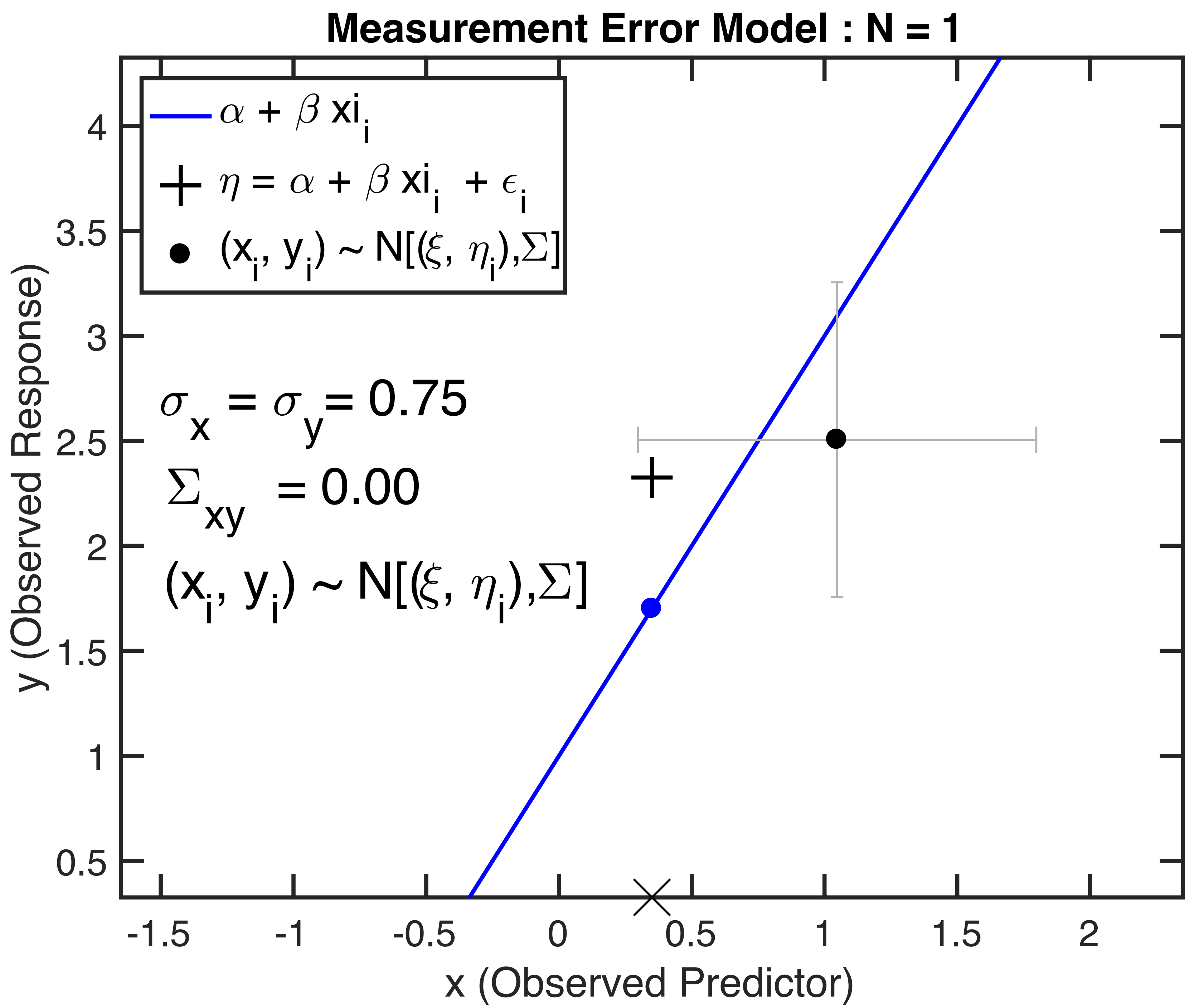
Step 1:
Generating Latent Variables from Parameters:

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

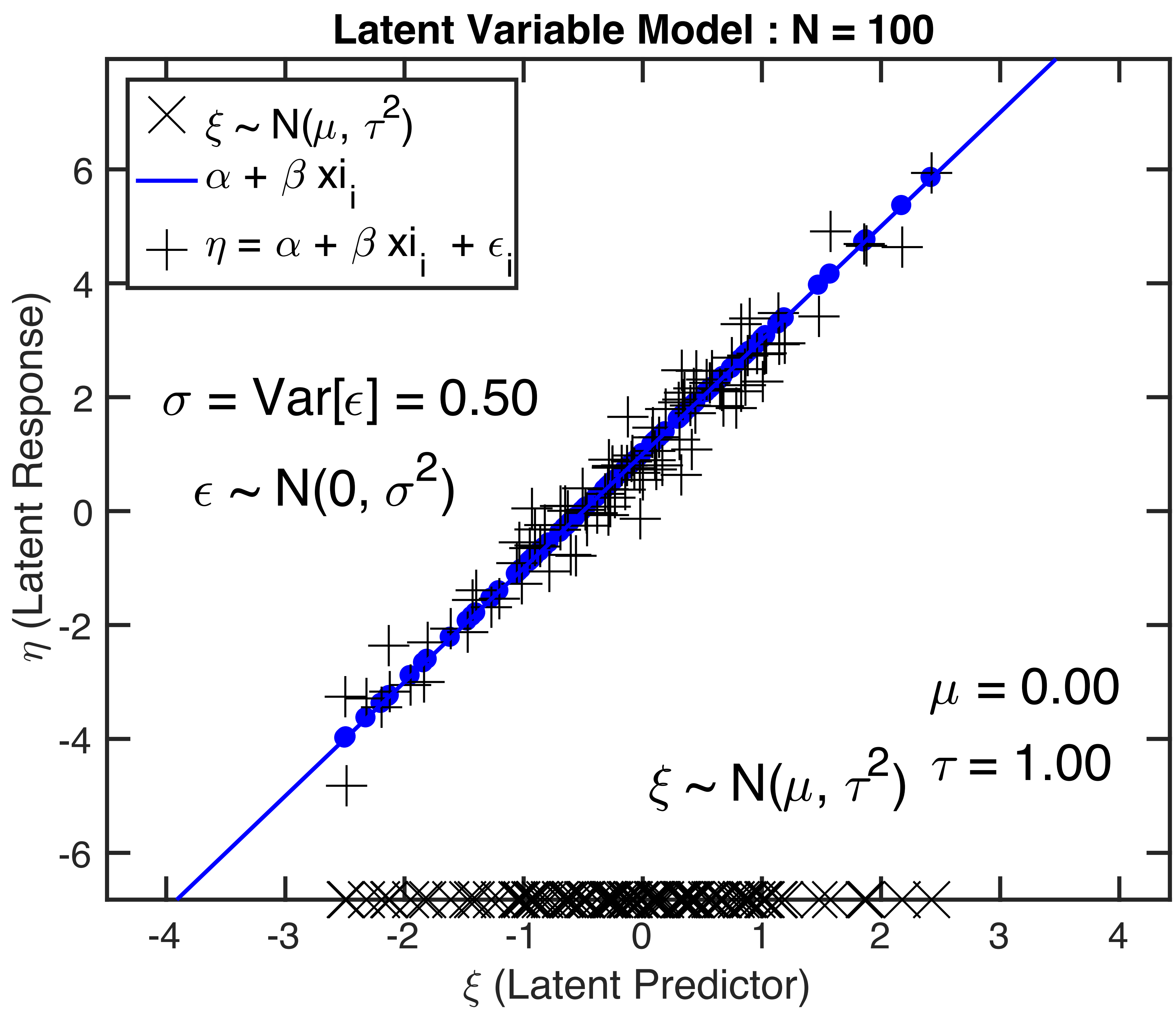


Step 2: Generating Observed Data from Latent Variables

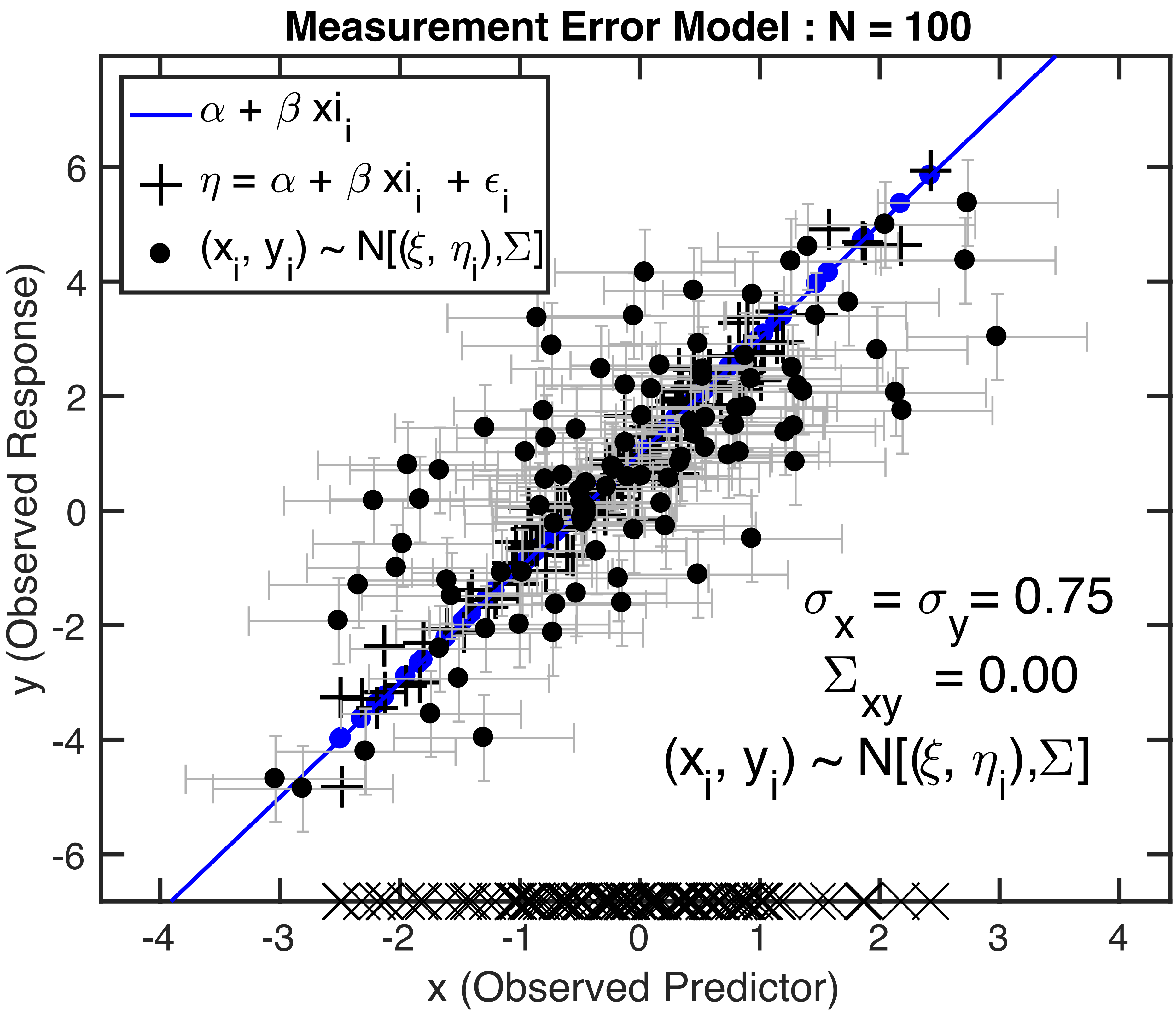
$$P([x_i, y_i] | \eta_i, \xi_i) = N([x_i, y_i] | [\eta_i, \xi_i], \mathbf{\Sigma})$$



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: Derive this!

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

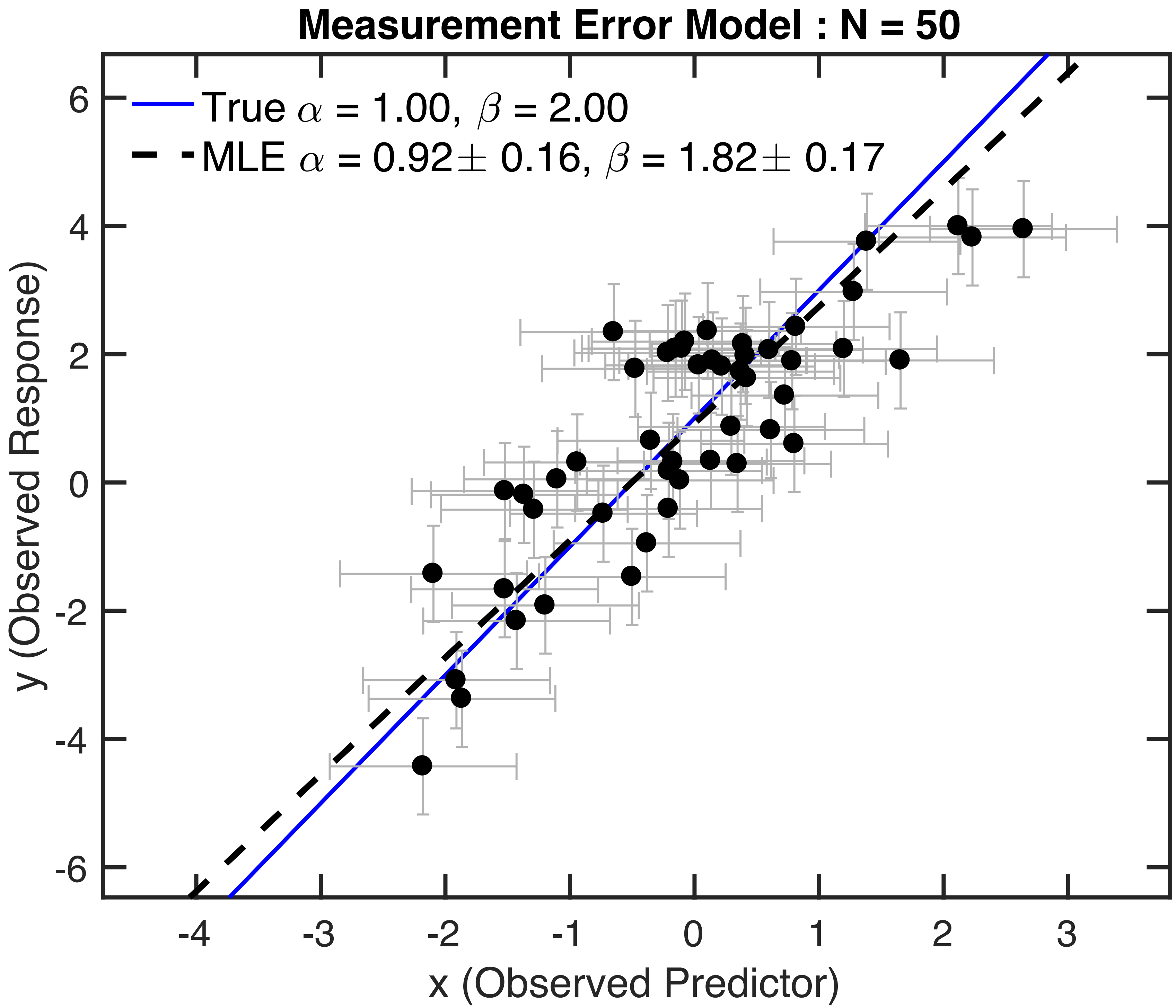
Gave More General Solution when P(ξ|Ψ)
is a Mixture of Gaussians
(set K=1, π_1 = 1 for us)

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

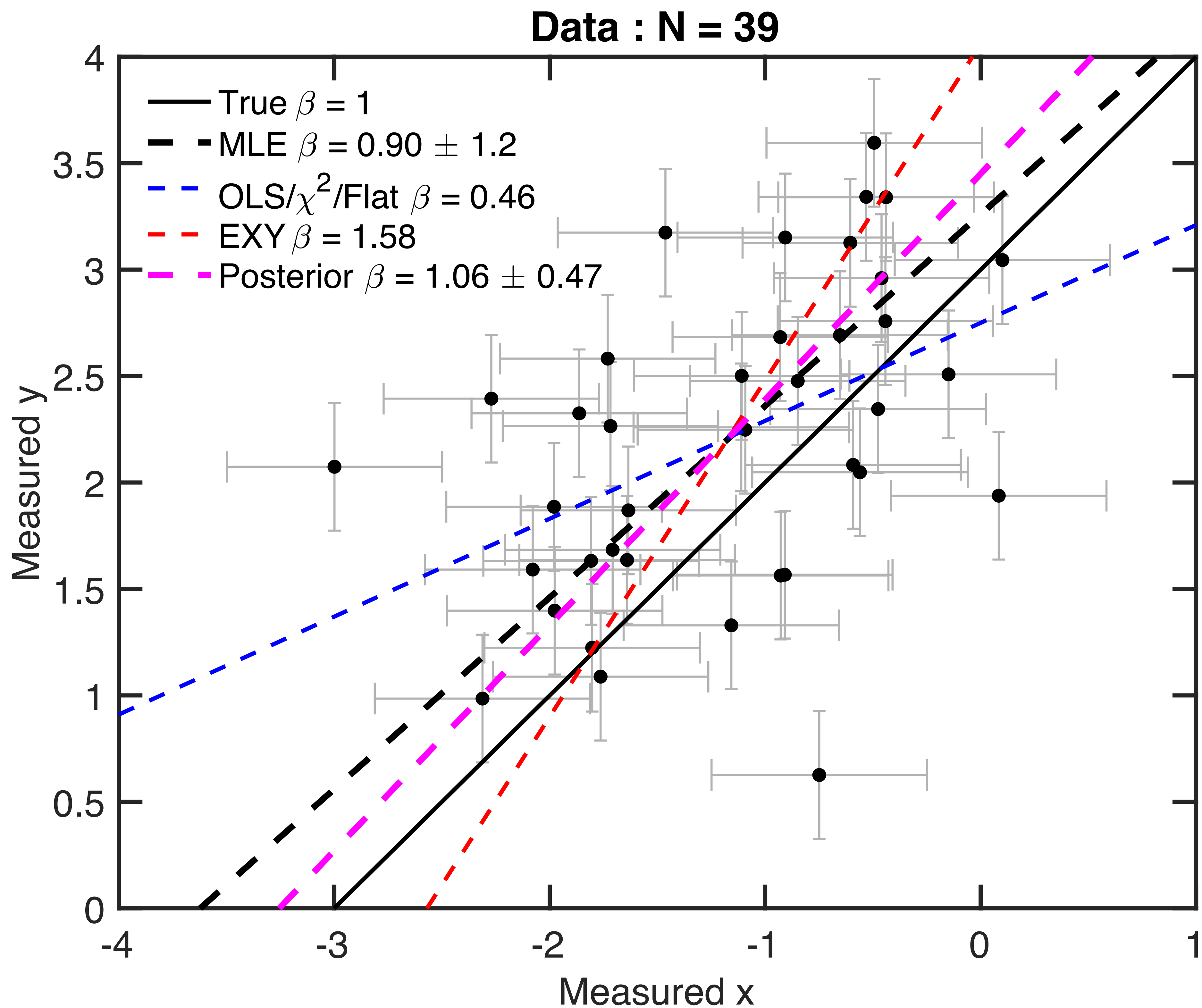
$$\boldsymbol{\zeta}_k = (\alpha + \beta\mu_k, \mu_k), \qquad (17)$$

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

Example

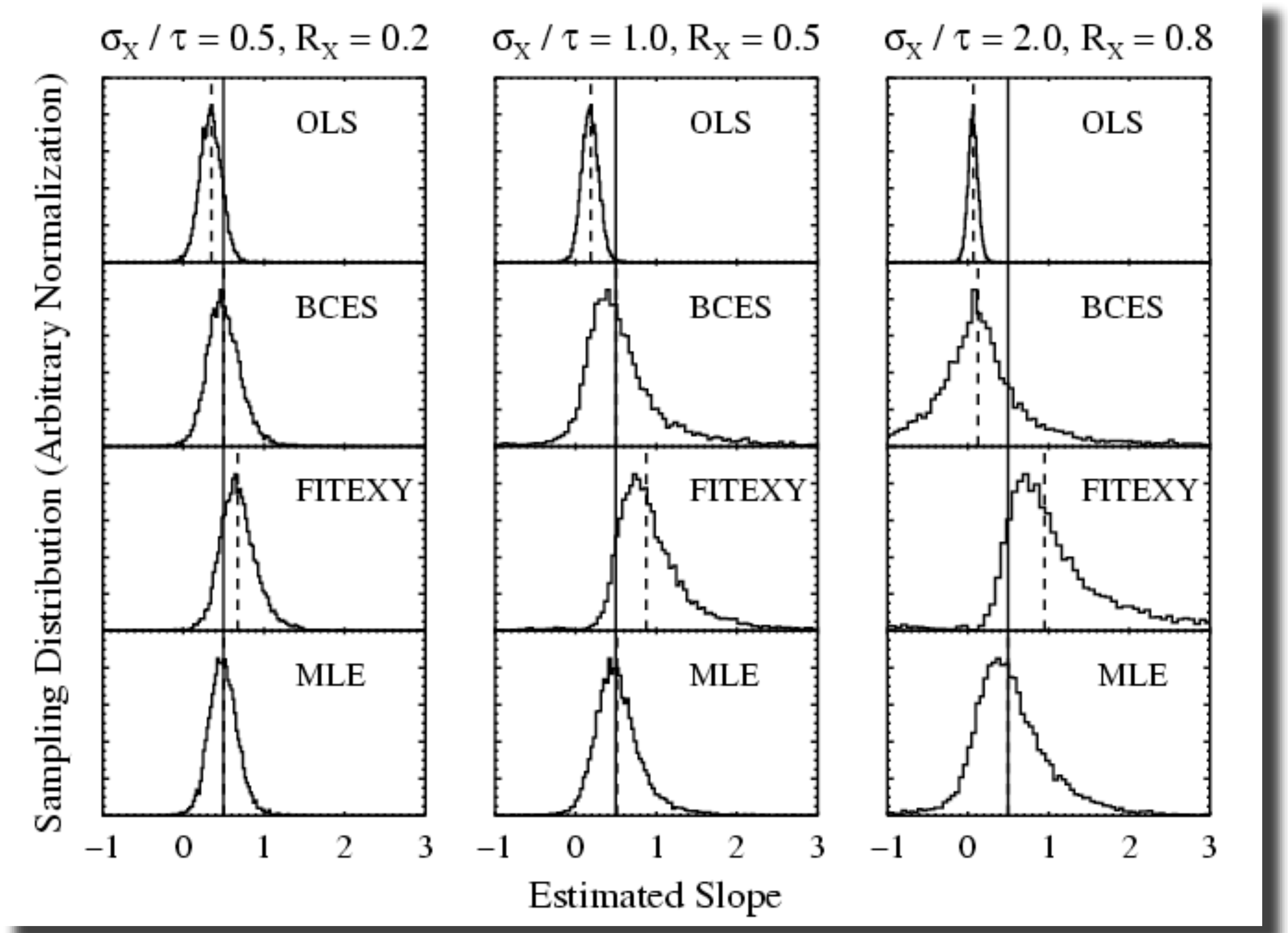

Measurement Error Model : N = 50
True α = 1.00, β = 2.00
MLE α = 0.92$\pm$ 0.16, β = 1.82$\pm$ 0.17
6
4
2
0
-2
-4
-6
-4
-3
-2
-1
0
1
2
3
y (Observed Response)
x (Observed Predictor)

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 $L(\theta)$? Derive it from the assumptions underlying your sampling distribution $P(D \mid \theta)$!
- Second question: what is your prior $P(\theta)$? (if Bayesian)
- Third question: How do I optimise/sample objective function to fit the data?

Bayesian Inference & Data Analysis

Recall: Probabilistic Generative Model for Linear Regression with (x,y) measurement errors

1. Population Distribution $\xi \sim N(\mu|\tau^2)$

Population Parameters: $\boldsymbol{\psi} = (\mu, \tau)$

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

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

Latent (true) Variables: (ξ_i, η_i)

3. Measurement Error: $[x_i, y_i] | \xi_i, \eta_i \sim N([\xi_i, \eta_i], \Sigma])$

Observed Data: (x_i, y_i)

Formulating Likelihood Function: Marginalising (integrating out) latent variables

"Complete Data Likelihood" (one datum)

$$P(x_i, y_i, \xi_i, \eta_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = \underbrace{P(x_i, y_i | \xi_i, \eta_i)}_{\text{Measurement Error}} \underbrace{P(\eta_i | \xi_i, \boldsymbol{\theta})}_{\text{Regression}} \underbrace{P(\xi_i | \boldsymbol{\psi})}_{\text{Population Distribution}}$$

"Observed Data Likelihood" (one datum): integrate out latent variables

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

Observed Data Likelihood (all data):

$$P(\boldsymbol{x}, \boldsymbol{y} | \, \boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} P(x_i, y_i | \, \boldsymbol{\theta}, \boldsymbol{\psi})$$

Knowns and Unknowns

Population Parameters: $\boldsymbol{\psi} = (\mu, \tau)$

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

Latent (true) Variables (ξ_i, η_i)

Observed Data: (x_i, y_i)

In Frequentist Statistics, distinction btw data and parameters: parameters are fixed and unknown, but not “random”. Only “data” are random realisations of random variables

Knowns and Unknowns

What is the nature of the latent variables (ξ_i, η_i) ?

They have a probability distribution:

$$(\xi_i, \eta_i) \sim P(\xi_i, \eta_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = P(\eta_i | \, \xi_i, \boldsymbol{\theta}) P(\xi_i | \, \boldsymbol{\psi})$$

Often called “nuisance parameters”
parameters you need to introduce to complete the model
but are not the parameters of interest: $(\boldsymbol{\theta}, \boldsymbol{\psi})$

Also referred to as “missing data”
quantities that you didn’t observe, but wish you had!
but relate to actual measurements (x,y)

Are the latent variables “data” or “parameters”?

Bayesian viewpoint

- There is a symmetry between data D and parameters θ - both are random variables described by probability distributions
- Actually they are described by a joint probability $P(D, \theta)$
- Data are random variables whose realisations are observed, parameters are RVs not observed
- Goal is to infer the unobserved parameters from the observed data using the rules of probability:
- Conditional Probability: $P(\theta| D) = P(D, \theta)/P(D)$
- Bayes' Theorem: $P(\theta| D) = P(D| \theta)P(\theta)/P(D)$
- Probability interpreted as degree of belief / uncertainty in hypotheses

Bayes' Theorem

Joint Probability of Data and Parameters:

$$P(D,\theta) = P(D|\,\theta)P(\theta) = P(\theta|\,D)P(D)$$

Probability of Parameters Given Data:

$$P(\theta|\,D) = P(D|\,\theta)P(\theta)/P(D)$$

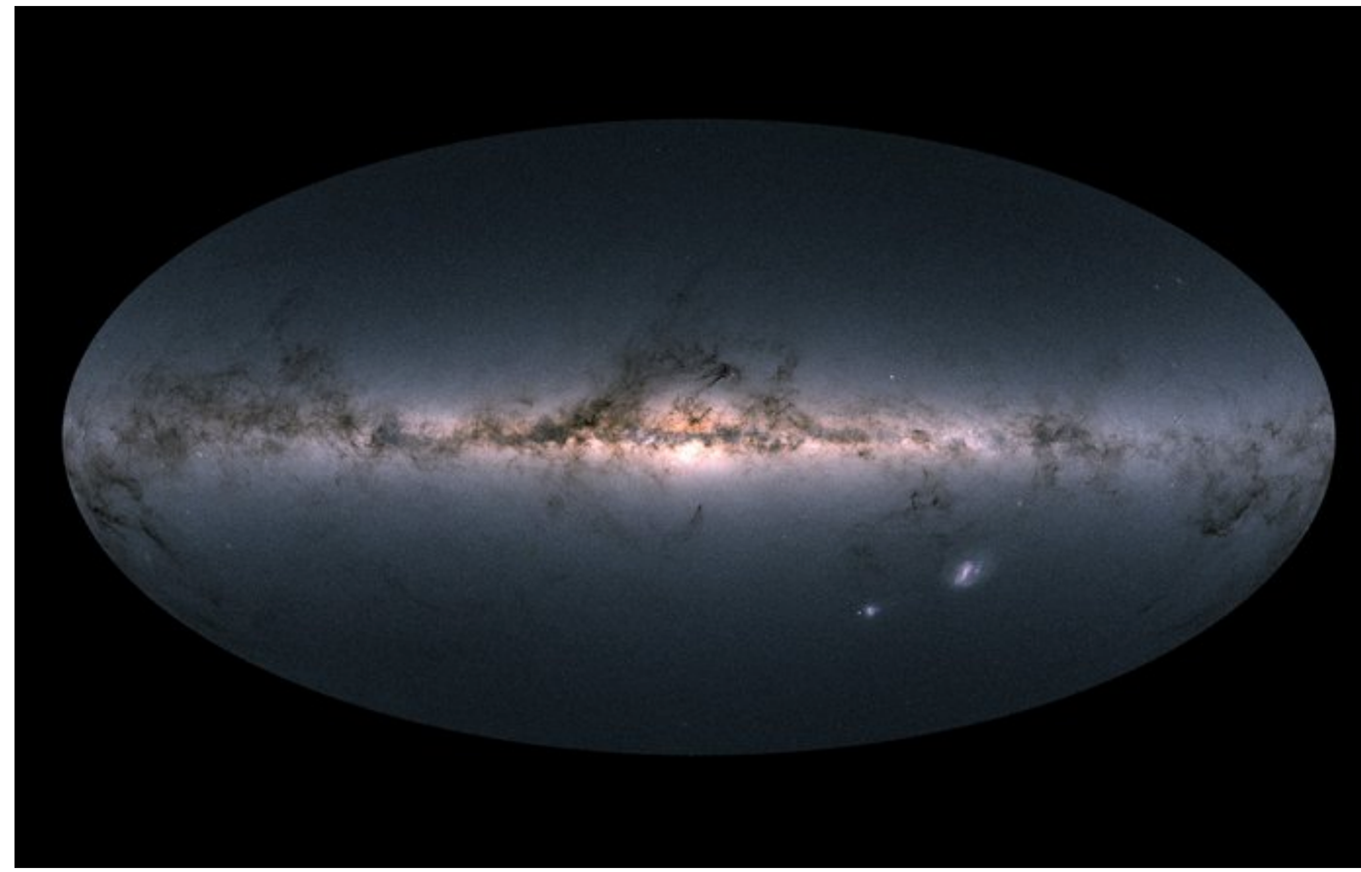
Normalisation Constant

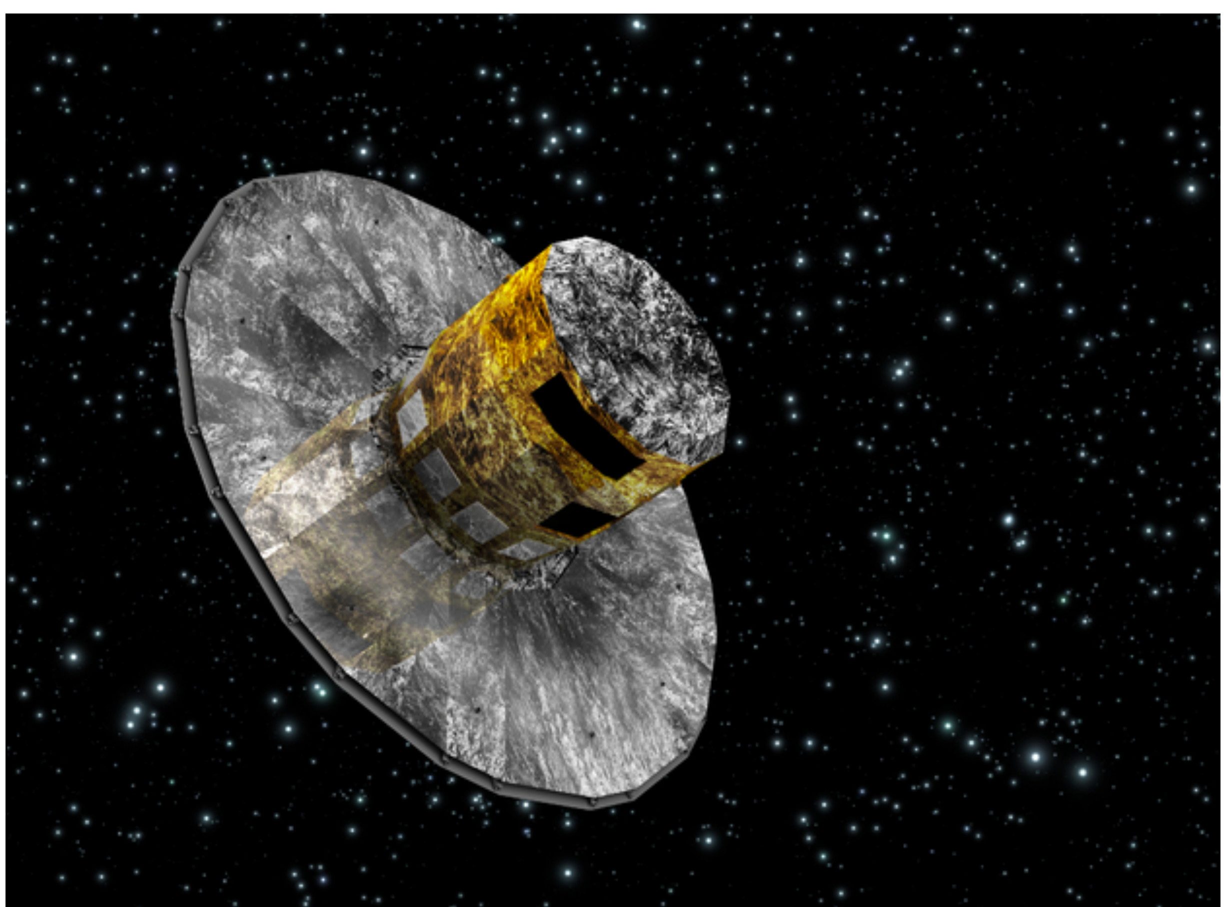
Posterior probability: Degree of Belief

Likelihood: Sampling Distribution

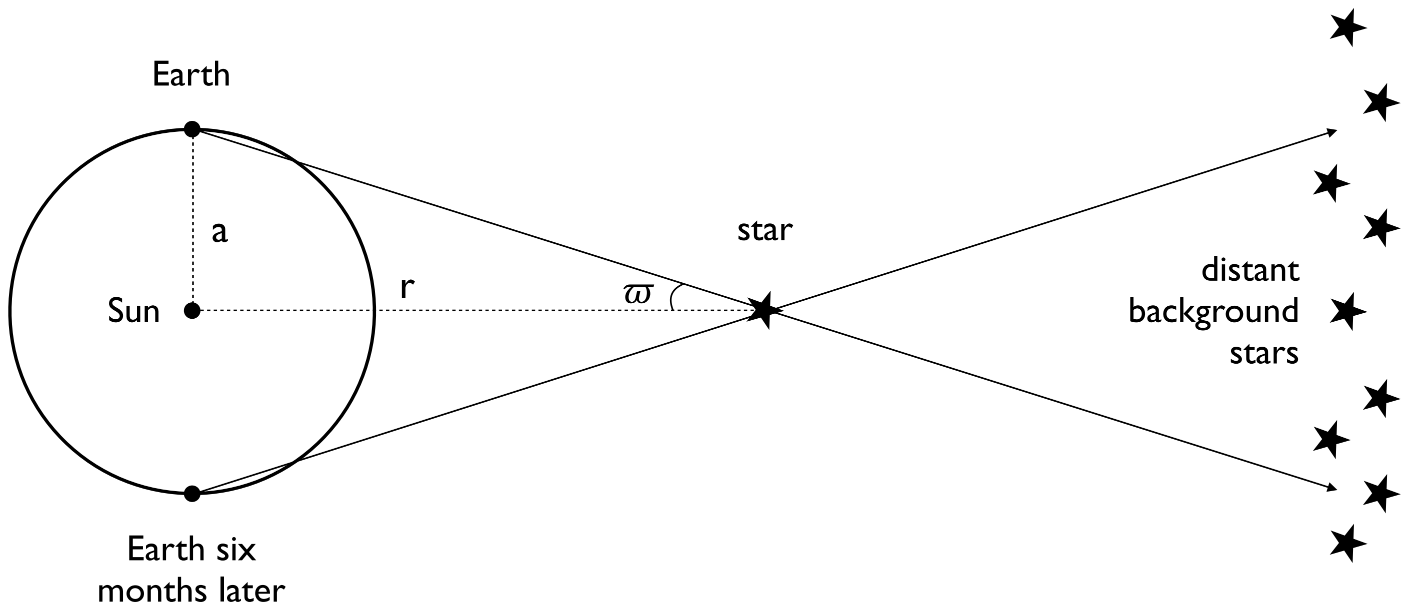
Prior probability: Degree of Belief

Gaia Parallaxes





Parallax Example



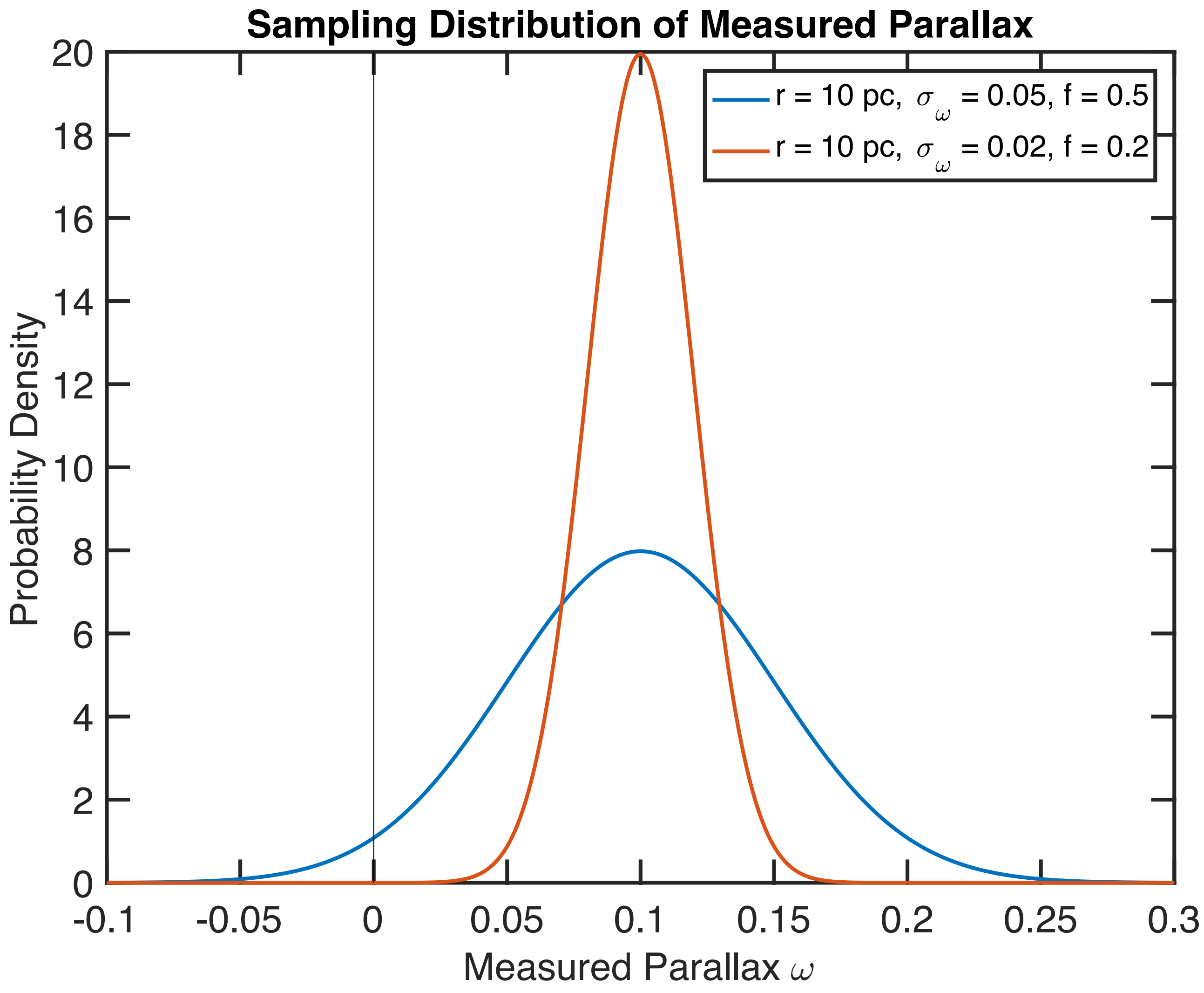
$$\frac{\omega}{\text{arcsec}} = \frac{\text{parsec}}{r}$$

The parallax ϖ of a star is the apparent angular displacement of that star (relative to distant background stars) due to the orbit of the Earth about the Sun. More precisely, the parallax is the angle subtended by the Earth's orbital radius a as seen from the star. As parallaxes are extremely small angles ($\varpi \ll 1$), $\varpi = a/r$ to a very good approximation. When ϖ is 1 arcsecond, r is defined as the *parsec*, which is about 3.1×10^{13} km. In this sketch the size of the Earth's orbit has been greatly exaggerated compared to the distance to the star, and the distance to the background stars in reality is orders of magnitude larger again.

C. Bailer-Jones."Estimating Distances from Parallaxes." 2015, PASP,127, 994, https://arxiv.org/abs/1507.02105

Parallax Measurement Error (Gaussian)

$$P(\varpi\,|\,r) \;=\; \frac{1}{\sigma_{\varpi}\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_{\varpi}^{2}}\left(\varpi-\frac{1}{r}\right)^{2}\right] \quad \text{where} \quad \sigma_{\varpi}>0,$$



- Negative parallax measurement possible due to observational noise
- Indicates that distance is likely to be large (parallax close to zero)
- Contains information

Inaccuracy of Gaussian Approx / Propagation of Error

$$P(\varpi\,|\,r) \;=\; \frac{1}{\sigma_\varpi\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi-\frac{1}{r}\right)^2\right] \quad \text{where} \quad \sigma_\varpi > 0,$$

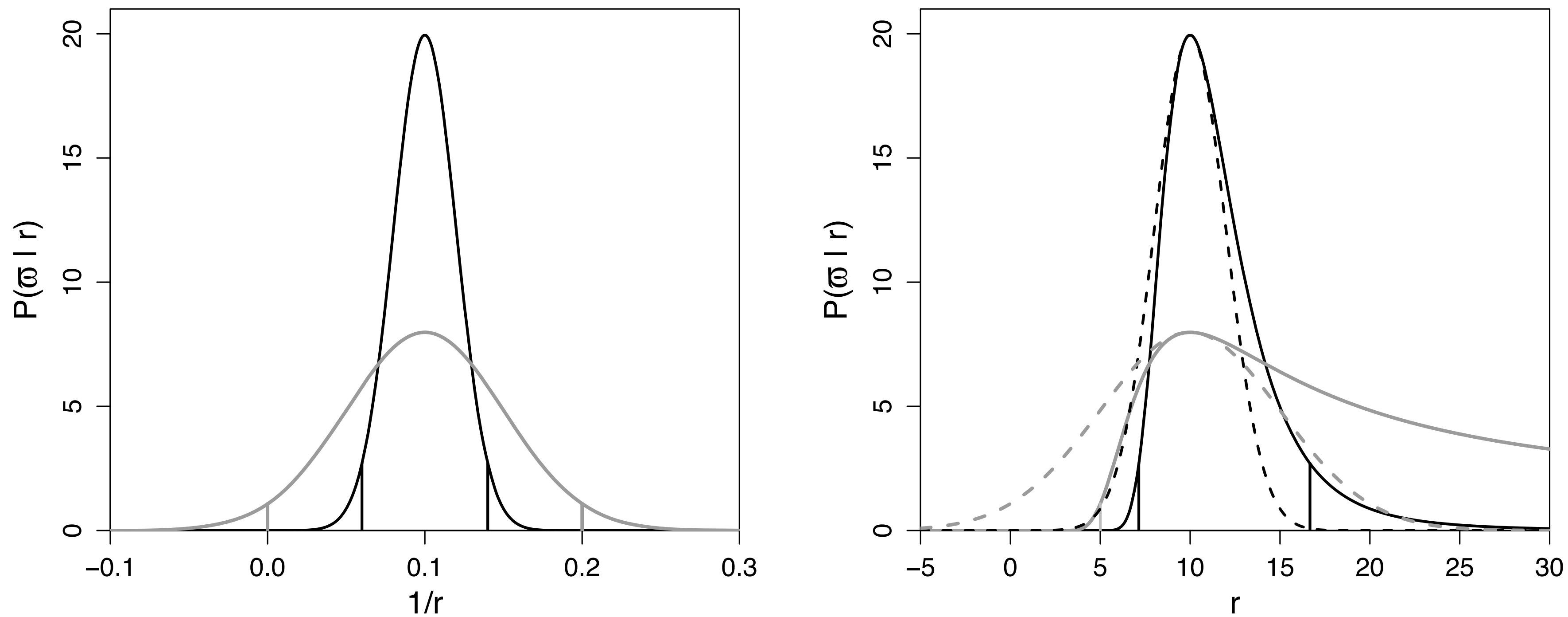
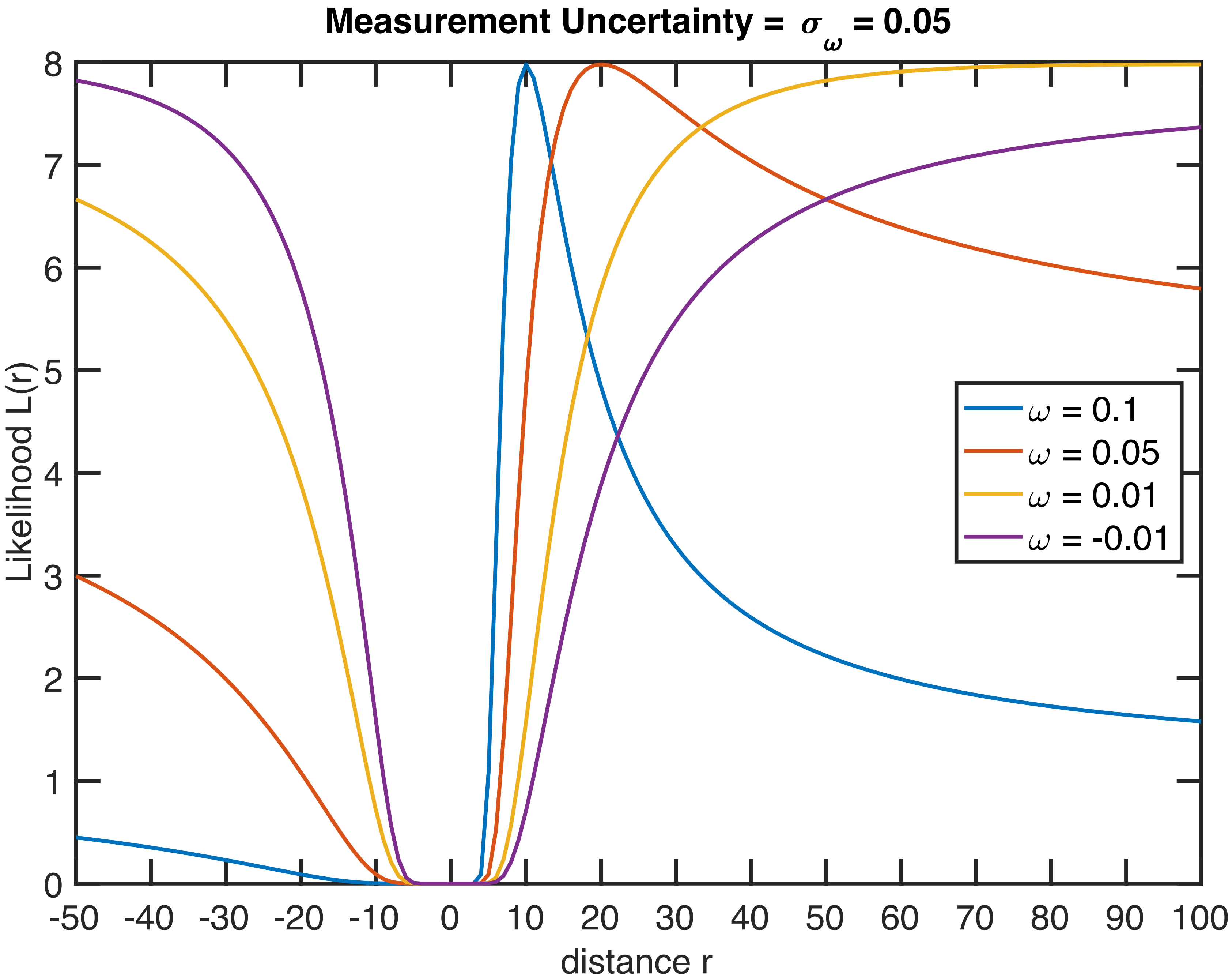


Fig. 3.4 The solid lines show the likelihood (equation 3.20) in the parallax problem for a measured parallax of $\varpi = 0.1$ as a function of $1/r$ in the left panel (a Gaussian) and as a function of r in the right panel. Note that the functions in the right panel are not PDFs over r, so they are not normalized. The black line is for $\sigma_\varpi = 0.02$ (so $f \equiv \sigma_\varpi/\varpi = 1/5$) and the grey line is for $\sigma_\varpi = 0.05$ ($f = 1/2$). The vertical lines denote the upper and lower 2σ limits around $1/\varpi$; the upper limit for the grey curve in the right panel is at $r = \infty$. The dashed lines in the right panel correspond to a Gaussian with mean $1/\varpi$ and standard deviation σ_ϖ/ϖ^2. Each of these Gaussians has been multiplied by the ratio of its standard deviation to that of the likelihood, $(\sigma_\varpi/\varpi^2)/\sigma_\varpi = 100$, in order to put them on the same vertical scale as the likelihood.

The Futility Function



- Likelihood is positive on negative values of distance
- Negative Measurements have no mode / MLE

An improper distance prior on r

Bayes' Thorem:

$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

$$P(r) = \begin{cases} 1 & r > 0 \\ 0 & \text{otherwise} \end{cases}$$

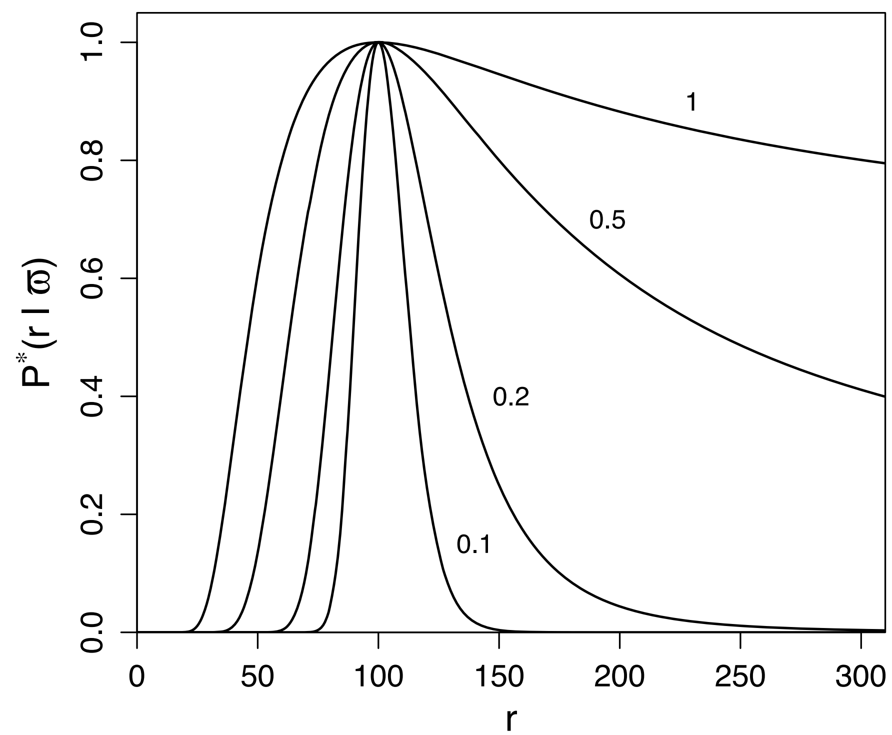
Require Positive Distances r

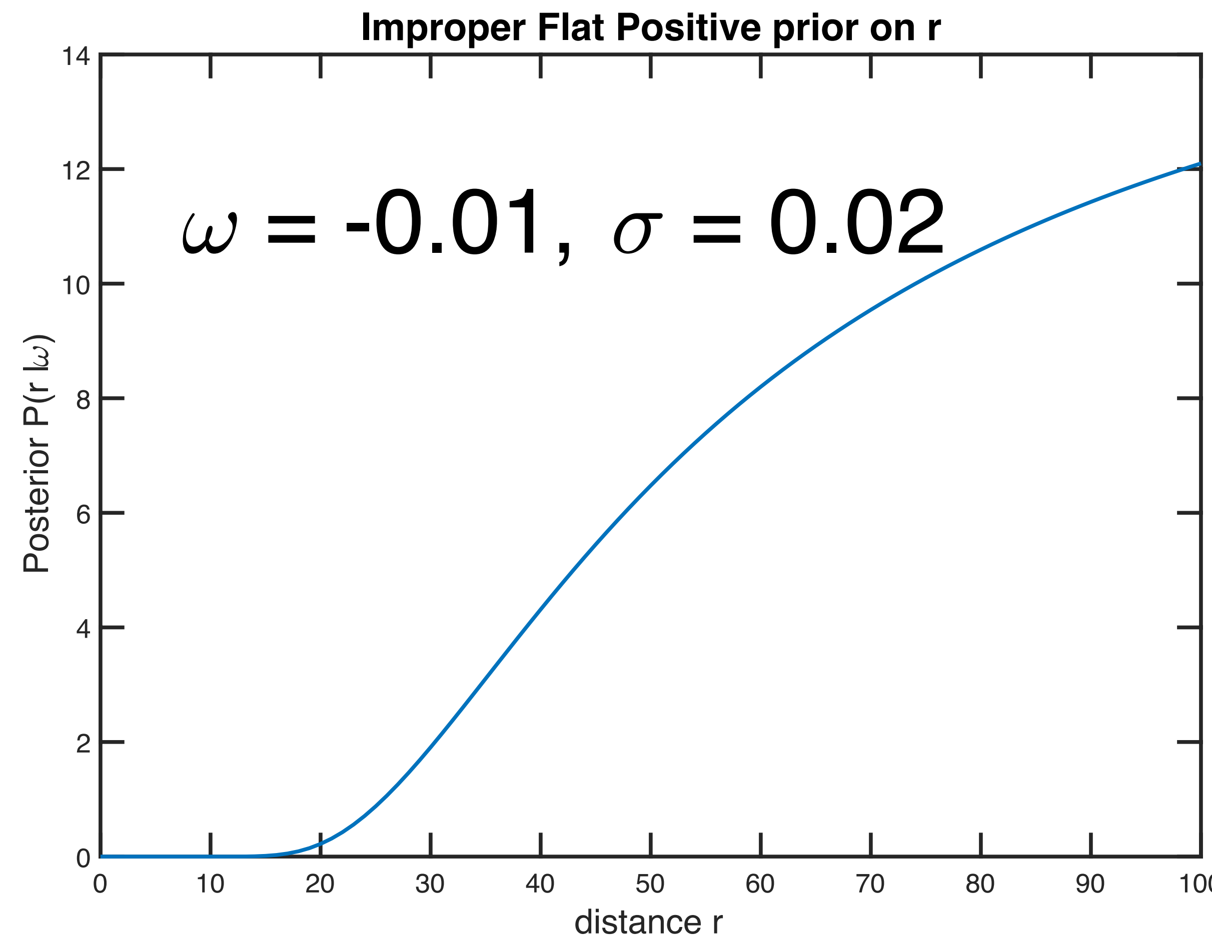
Improper = not normalisable

Posterior of r

$$\omega = 0.01$$

$$f = \frac{\sigma_\omega}{\omega} = 0.1,\ 0.2,\ 0.5,\ 1.0$$





The unnormalized posterior using the improper uniform prior (equation 3.25) for $\varpi = 1/100$ and four values of $f = 0.1, 0.2, 0.5, 1.0$. The unnormalized posteriors have been scaled to all have their mode at $P^*(r \,|\, \varpi) = 1$. Figure reproduced from Bailer-Jones (2015).

Improper Posterior: No mean, variance, etc.

Mode (1/$\boldsymbol{\omega}$) exists for positive $\boldsymbol{\omega}$

but doesn't for negative $\boldsymbol{\omega}$

Proper Distance prior

$$P(r) = \begin{cases} \dfrac{1}{r_{\rm lim}} & \text{if } \ 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases}$$

r_{lim} = maximum possible distance of a star in our survey

$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

Proper Distance prior

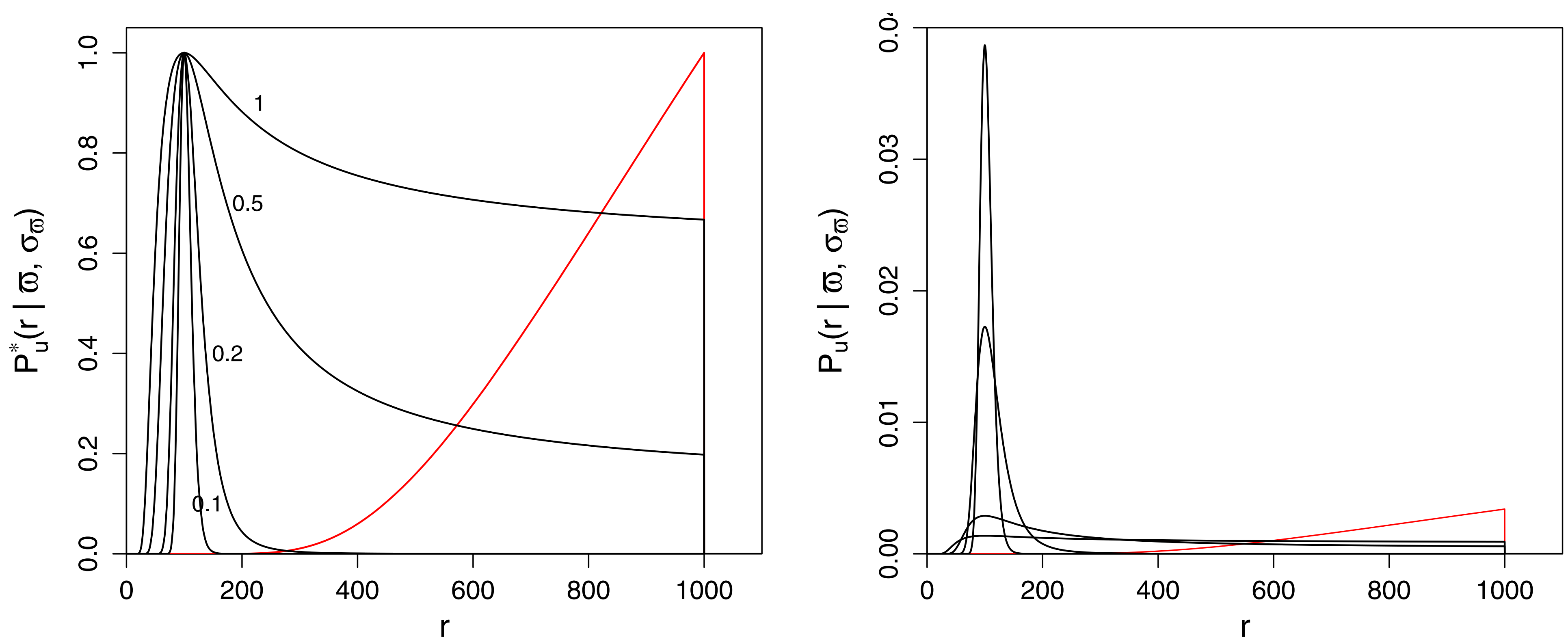


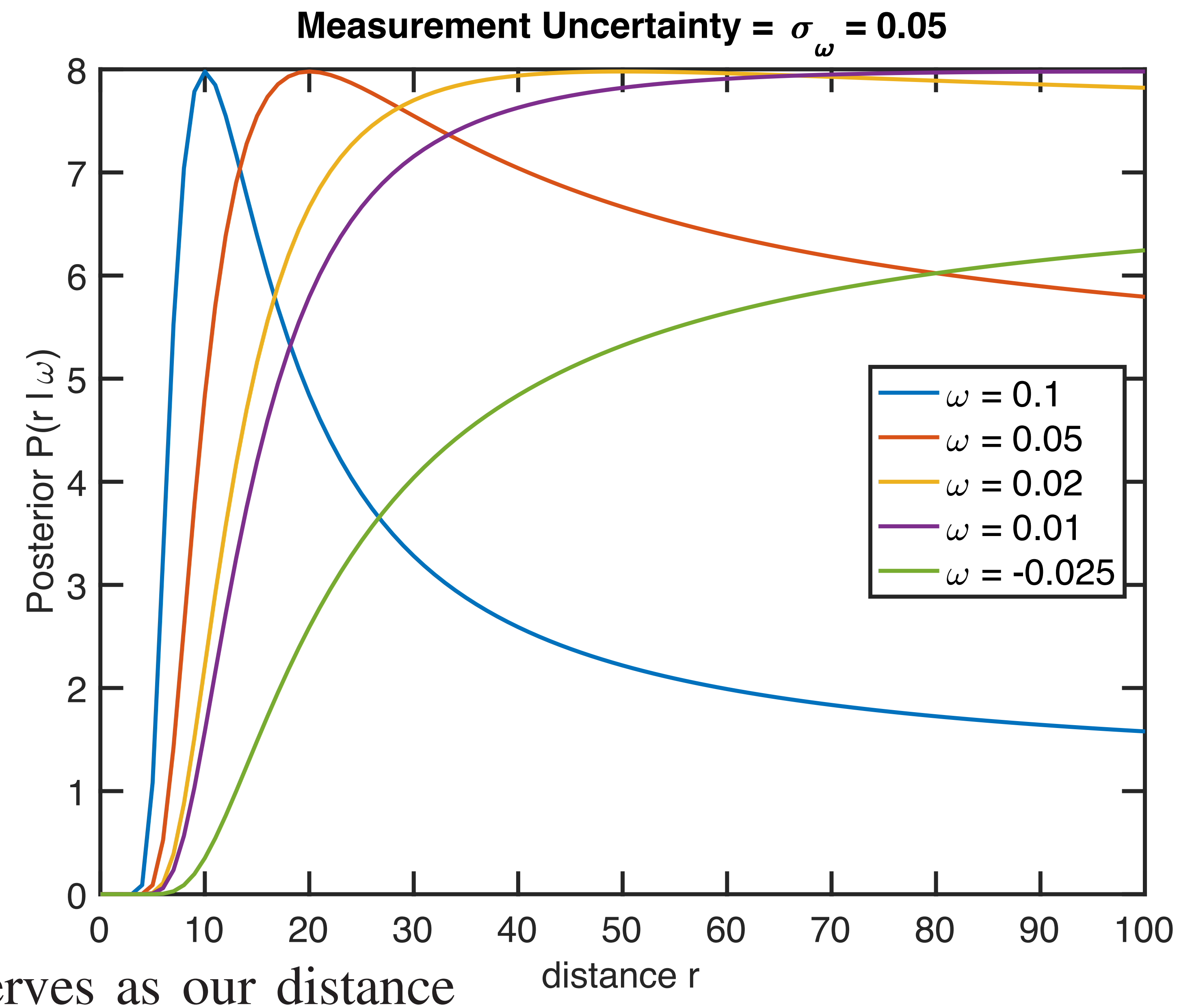
FIG. 4.—Left: the unnormalized posterior $P_u^*(r|\varpi, \sigma_\varpi)$ (truncated uniform prior with $r_{\rm lim} = 10^3$) for $\varpi = 1/100$ and four values of $f = (0.1, 0.2, 0.5, 1.0)$ (*black lines*). The red line shows the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The posteriors have been scaled to all have their mode at $P_u^*(r|\varpi, \sigma_\varpi) = 1$. Right: the same five posterior PFDs but now normalized. See the electronic edition of the *PASP* for a color version of this figure.

Normalisable posterior works for negative **ω**

“Pile-up” at upper limit for small measured parallax (**ω** < σ)

Proper Uniform Distance prior

$r_{\mathrm{lim}} = 100$



(Stellar density ~ $1/r^2$)

The mode of the resulting posterior serves as our distance estimator, and is

$$r_{\mathrm{est}} = \begin{cases} \frac{1}{\varpi} & \text{if } 0 < \frac{1}{\varpi} \le r_{\mathrm{lim}} \\ r_{\mathrm{lim}} & \text{if } \frac{1}{\varpi} > r_{\mathrm{lim}} \\ r_{\mathrm{lim}} & \text{if } \varpi \le 0 \end{cases} \quad \text{(extreme mode)}\,. \qquad (12)$$

"Pile-up" of modes at edge r_{lim} for small measured parallax ($\boldsymbol{\omega} < \sigma$)

Introducing physical constraints into the prior

$$P(r) = \begin{cases} \dfrac{3}{r_{\rm lim}^3}\, r^2 & \text{if } 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases}$$

Implies a uniform volume density of stars
up to maximum $r_{\rm lim}$

$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

$$P^*_{r^2}(r|\varpi, \sigma_\varpi) = \begin{cases} \frac{r^2}{\sigma_\varpi} \exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{r}\right)^2\right] & \text{if } 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases}.$$

Introducing physical constraints into the prior: Uniform volume density of stars

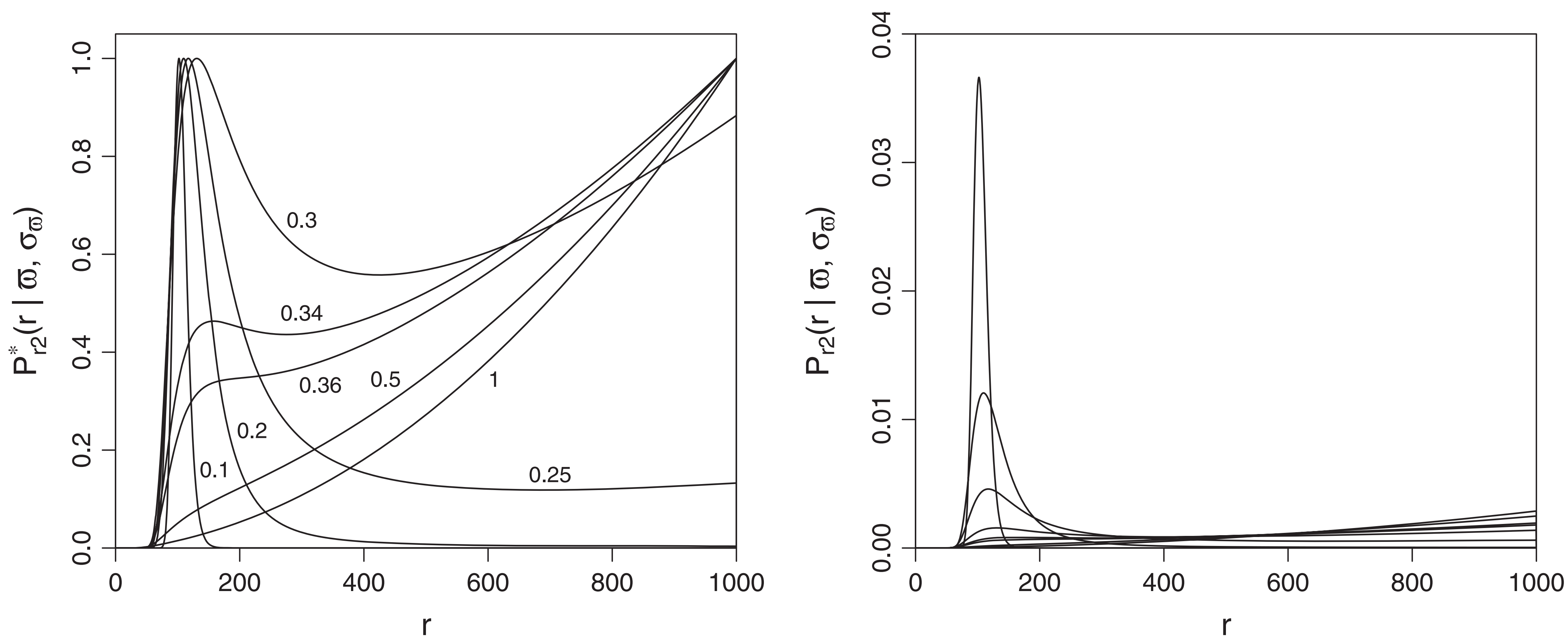


FIG. 7.—Left: the unnormalized posterior $P^*_{r^2}(r|\varpi,\sigma_\varpi)$ (truncated constant volume density prior with $r_{\rm lim}=10^3$) for $\varpi=1/100$ and eight values of $f=(0.1,0.2,0.25,0.3,0.34,0.36,0.5,1.0)$. The posteriors have been scaled to all have their mode at $P^*_{r^2}(r|\varpi,\sigma_\varpi)=1$. Right: the same posterior PDFs but now normalized. The curves with the clear maxima around $r=100$ are $f=(0.1,0.2,0.25,0.3)$ in decreasing order of the height of the maximum.

Introducing physical constraints into the prior: Uniform volume density of stars

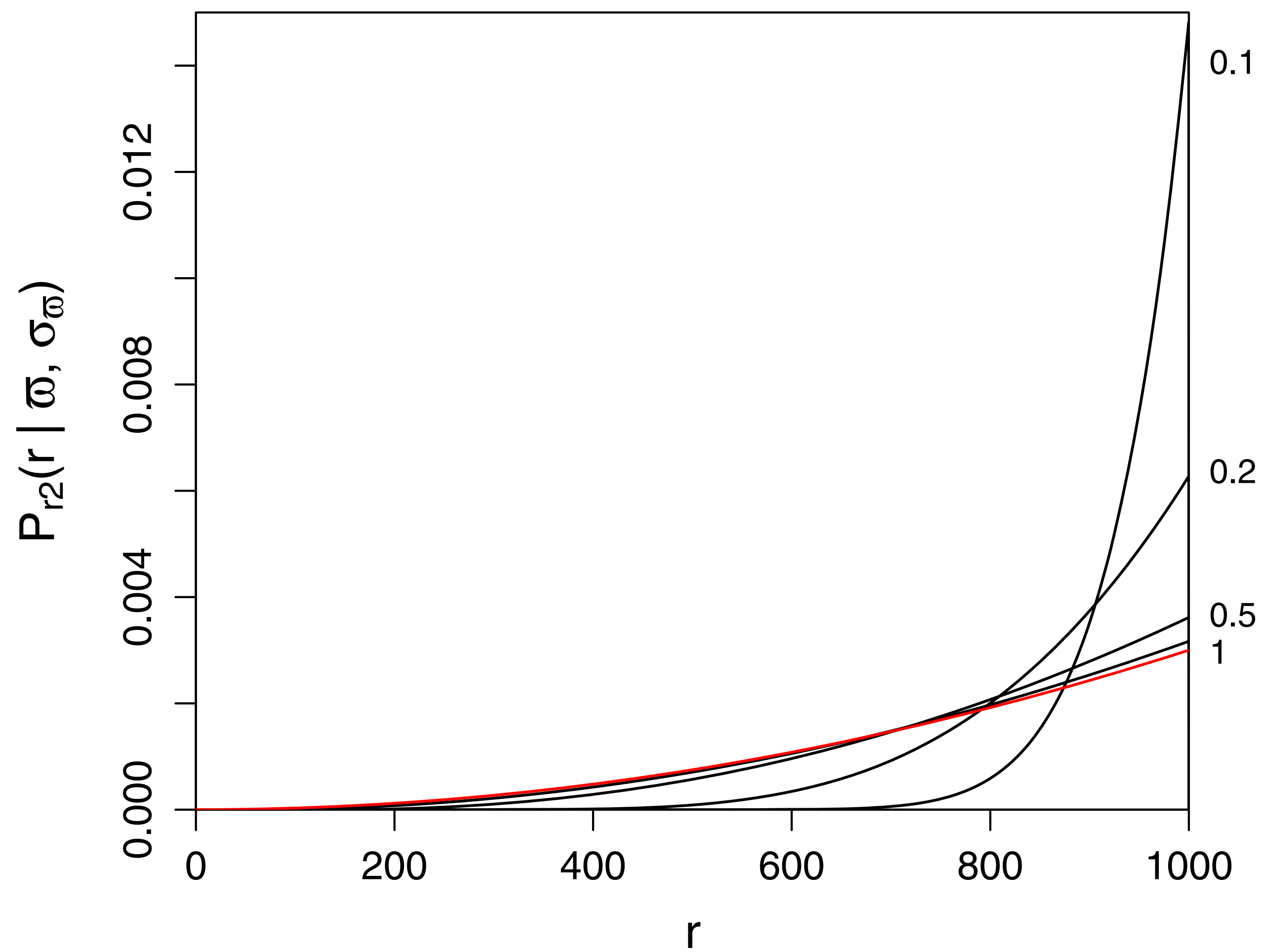


FIG. 9.—The normalized posterior $P^*_{r^2}(r|\varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\rm lim} = 10^3$) for $\varpi = -1/100$ and four values of $|f| = (0.1, 0.2, 0.5, 1.0)$ (*black lines*). The red line shows the posterior for $\varpi = 0$ for $\sigma_\varpi \gg 1/r_{\rm lim}$ (f is then undefined). See the electronic edition of the *PASP* for a color version of this figure.

Works for negative and zero parallaxes $\boldsymbol{\omega}$

but again pile-up at hard limit

Introducing physical constraints into the prior

$$P(r) = \begin{cases} \dfrac{1}{2L^3}\, r^2 e^{-r/L} & \text{if } \; r > 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential decrease in stars with Galactic length scale L

$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

$$P^*_{r^2 e^{-r}}(r|\varpi, \sigma_\varpi) = \begin{cases} \frac{r^2 e^{-r/L}}{\sigma_\varpi} \exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{r}\right)^2\right] & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Introducing physical constraints into the prior

Exponential decrease in stars with Galactic length scale L

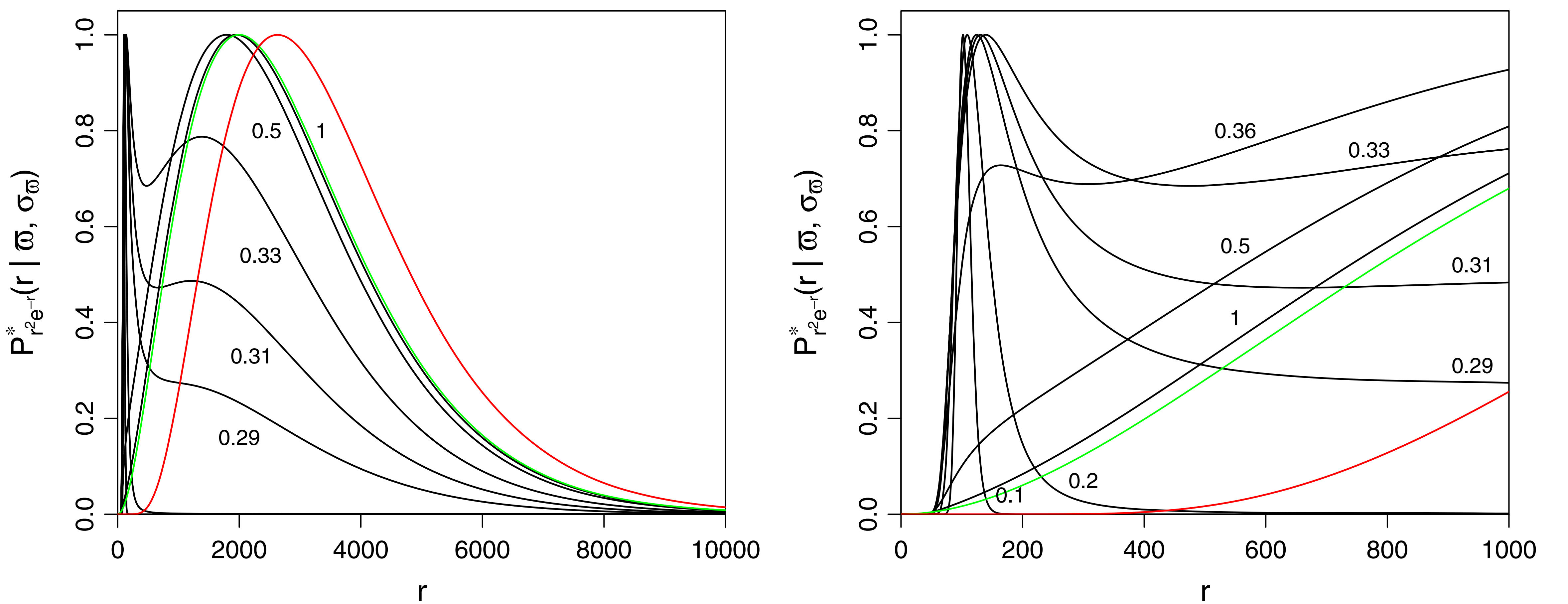


FIG. 12.—The *black lines* in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The *red line* is the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The *green curve* is the prior. The right panel is a zoom of the left one and also shows an additional posterior for $f = 0.36$. All curves have been scaled to have their highest mode at $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi) = 1$ (outside the range for some curves in the right panel). See the electronic edition of the *PASP* for a color version of this figure.

Smooth solutions : no hard edge in prior