Brandon C. Kelly
Hubble Fellow
Harvard-Smithsonian Center for Astrophysics

Measurement Error Models in Astronomy
The Data Collection Process

Astrophysical Process

Detector Collects Photons, Adds Noise

Random Number of Photons Reach our Detector

Need to use observed, contaminated data to draw conclusions about astrophysical sources, populations

Data Reduction, calibration add additional uncertainty

Brandon C. Kelly, bckelly@cfa.harvard.edu
No Intrinsic Scatter

- Random error about model is only due to measurement error
- Typical when analyzing a single observation from a single object
  - Example: Fitting an astrophysical model to a spectrum
- Methods are well-established for Gaussian Data (e.g., $\chi^2$, Bevington 2003)
- For count data, use the Poisson likelihood function (e.g., van Dyke 2001)
- Methods for incorporating calibration uncertainties are only recently being developed (e.g., Lee et al. 2011)

Shen et al. (2008)
Random variations due to both measurement error and source-to-source differences in physical properties

- Example: Correlations between supermassive black hole mass and host galaxy properties

- Typical when analyzing astronomical populations

- Often called ‘error in variables model with equation error’ in statistics

Gültekin et al. (2009)
Additive Measurement Error Model for Regression with Intrinsic Scatter

\[ \eta_i = f(\xi_i, \theta) + \varepsilon_i, \quad i = 1, \ldots, n, \quad E(\varepsilon_i) = 0, \quad E(\varepsilon_i^2) = \sigma^2 \]

\[ y_i = \eta_i + \varepsilon_{y,i}, \quad E(\varepsilon_{y,i}) = 0, \quad E(\varepsilon_{y,i}^2) = \sigma_{y,i}^2 \]

\[ x_i = \xi_i + \varepsilon_{x,i}, \quad E(\varepsilon_{x,i}) = 0, \quad E(\varepsilon_{x,i}^T \varepsilon_{x,i}) = \Sigma_{x,i} \]

- \( \eta_i, \xi_i \): True value for dependent and independent variable for the \( i \)th data point (i.e., the response and covariate)
- \( y_i, x_i \): Measured value for dependent and independent variable
- \( \theta \): Parameters for function that describes how mean \( \eta \) depends on \( \xi \)

Example: For linear regression,

\[ f(\xi, \theta) = \alpha + \xi^T \beta, \quad \theta = (\alpha, \beta, \sigma) \]
Measurement Error Results in a Loss of Information

- Under the additive model, measured distribution is true distribution convolved with measurement error distribution
- Blurs the observed distribution (like a PSF)
  - Loss of fine structure
  - Trends are less apparent, appear to be weaker
How Does Meas. Err. Effect Linear Regression?

\[
\hat{\beta} = \frac{\text{Cov}(x,y)}{\text{Var}(x)} = \frac{\text{Cov}(\xi,\eta)}{\text{Var}(\xi) + \sigma_x^2}
\]

- Trend appears flatter

\[
\hat{\sigma}^2 = \text{Var}(y - \alpha - \hat{\beta}x)
= (\hat{\beta}^2 - \hat{\beta}^2)\text{Var}(\xi) + \hat{\beta}^2 \sigma_x^2 + \sigma_y^2 + \sigma^2
\]

- Intrinsic Dispersion appears larger

\[
\text{Var}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\text{Var}(x)} = \frac{\hat{\sigma}^2}{\text{Var}(\xi) + \sigma_x^2}
\]

- Estimated uncertainty in slope goes to zero when \(\sigma_x\) becomes large

6/15/11 Brandon C. Kelly, bckelly@cfa.harvard.edu
Functional vs. Structural Models

- **Functional Models:**
  - True values of variables are considered to be unknown but fixed
  - Astro examples include BCES estimator (Akritas & Bershady 1996) and FITEXY (Press et al. 2007)

- **Structural Models:**
  - True values of variables are additional parameters to be estimated or marginalized over
  - Need to assume a distribution for true values of variables
  - Examples include maximum-likelihood estimators and Bayesian methods
Method of Moments (e.g., Fuller 1987)
- Least-squares estimate of slope(s) is derived from moments of distribution
- Correct Moments of measured data for measurement error

Modified loss functions (e.g., Press et al. 2007)
- Loss function (i.e., figure of merit function, e.g., $\chi^2$) typically assumes no measurement error
- Modify loss function to include measurement error

Neither method assumes a particular distribution for the independent variable(s), the intrinsic scatter, or the measurement errors
Method of Moments (MoM)

MoM estimators are asymptotically unbiased and normally distributed

Method of Moments (MoM) Estimators are asymptotically unbiased and normally distributed.


Ordinary Least Squares (OLS) Estimators

\[
\hat{\beta}_{\text{OLS}} = \frac{\text{Cov}(x, y)}{\text{Var}(x)}
\]

\[
\hat{\alpha}_{\text{OLS}} = y - \hat{\beta}_{\text{OLS}} x
\]

\[
\hat{\sigma}_{\text{OLS}}^2 = \text{Var}(y) - \hat{\beta}_{\text{OLS}} \text{Cov}(x, y)
\]

\[
\hat{\sigma}_{x}^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_{x, i}^2
\]

\[
\hat{\sigma}_{y}^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_{y, i}^2
\]
Most common loss function in astronomy: Weighted squared error (i.e., $\chi^2$):

$$Q(\theta) = \sum_{i=1}^{n} \frac{(y_i - f(x_i, \theta))^2}{\sigma_{y,i}^2}$$

Estimate $\theta$ by minimizing $Q(\theta)$

For linear regression with no intrinsic scatter, can modify $Q(\theta)$ to be (Sprent 1966):

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \frac{(y_i - \alpha - \beta x_i)^2}{\sigma_{y,i}^2 + \beta^2 \sigma_{x,i}^2}$$
For intrinsic scatter, a number of authors have suggested (e.g., Tremaine et al. 2002):

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

But this estimator has some undesirable qualities:
- Minimum occurs when $\sigma$ goes to $\infty$
- Statistical properties (e.g., bias and variance) unknown
- Simulations by Kelly (2007) suggest this estimator is biased
Instead of reweighting $Q(\alpha, \beta, \sigma)$, try subtracting off contribution from meas err:

$$Q'(\alpha, \beta, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^{n} [(y_i - \alpha - \beta x_i)^2 - \sigma_{y,i}^2 - \beta^2 \sigma_{x,i}^2]$$

Minimizer of $Q'(\alpha, \beta, \sigma)$ is the Method of Moments Estimator
Structural Models for Regression and Density Estimation

- Need to assume a parameteric form for distribution of independent variable(s), intrinsic scatter, and measurement errors
- Includes maximum-likelihood estimators and Bayesian techniques
- Typically produce less variable (i.e., more precise) estimates, more accurate uncertainties, compared to functional models
The Measured Data Likelihood Function

- Both maximum-likelihood estimators and Bayesian methods require a likelihood function
- Denote $\psi$ to be set of parameters for distribution of true independent variable
- Measured data likelihood function:

$$p(y_i, x_i | \theta) = \iint p(y_i, x_i, \eta_i, \xi_i | \theta, \psi) d\eta_i d\xi_i$$
Deriving the Measured Data Likelihood Function

| Intrinsic distribution of independent variable(s) |
| Distribution of dependent variable at fixed independent variable(s) |
| Measurement error distribution |

\[ \xi_i \mid \psi \sim p(\xi_i \mid \psi) \]

\[ \eta_i \mid \xi_i, \theta \sim p(\eta_i \mid \xi_i, \theta) \]

\[ y_i, x_i \mid \eta_i, \xi_i \sim p(y_i, x_i \mid \eta_i, \xi_i) \]

Measured Data Likelihood Function is:

\[
p(y_i, x_i \mid \theta) = \iint p(y_i, x_i, \eta_i, \xi_i \mid \theta, \psi) d\eta_i d\xi_i
\]

\[= \iint p(y_i \mid \eta_i) p(x_i \mid \xi_i) p(\eta_i \mid \xi_i, \theta) p(\xi_i \mid \psi) d\eta_i d\xi_i\]
Assume a prior distribution, $p(\theta, \psi)$, to convert likelihood function to probability of parameters, given measured data (posterior):

$$p(\theta, \psi \mid y_i, \ldots, y_n, x_1, \ldots, x_n) \propto p(\theta, \psi)p(y_i, \ldots, y_n, x_1, \ldots, x_n \mid \theta, \psi)$$
Comparison of Bayesian, BCES (MoM), and Modified Loss Function

Kelly (2007)

PME = Posterior Median Estimate under Gaussian Model
For many applications, integrating out the missing true values is not analytically possible.

Instead, MCMC provides a means of simulating random draws of both the parameters and the missing data.

- No need to do integration
- Also get estimates of true values for data set
- Often simpler to implement than maximum-likelihood
MCMC for Measurement Error Models

- **Step 1:** Propose new values for true values, given current parameters and measured values

  \[ \eta_i^{new} \sim p(\eta_i | \xi_i, \theta, y_i), \quad \xi_i^{new} \sim p(\xi_i | \eta_i, \psi, x_i) \]

- **Step 2:** Propose new values for regression parameters, given current true values

  \[ \theta^{new} \sim p(\theta | \eta_1, \ldots, \eta_n, \xi_1, \ldots, \xi_n) \]

- **Step 3:** Propose new values for \( \psi \), given current true values

  \[ \psi^{new} \sim p(\psi | \xi_1, \ldots, \xi_n) \]
Example: Bayesian Segmented Regression

Constantin+, in prep

Posterior Mean for true values
Example: Bayesian Segmented Regression

Posterior Mean for true values
# Summary: Comparison of Structural and Functional Models

<table>
<thead>
<tr>
<th>FUNCTIONAL</th>
<th>STRUCTURAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Includes Methods of Moments (BCES)</td>
<td>▪ Includes maximum-likelihood and Bayesian methods</td>
</tr>
<tr>
<td>▪ Advantages</td>
<td>▪ Advantages</td>
</tr>
<tr>
<td>▪ Do not need to assume a distribution for true value of ind. var., intrinsic scatter, or measurement errors</td>
<td>▪ More stable, higher precision</td>
</tr>
<tr>
<td>▪ Computationally fast</td>
<td>▪ More accurate uncertainties</td>
</tr>
<tr>
<td>▪ Disadvantages</td>
<td>▪ MCMC enables Bayesian methods to be applied to highly complex problems</td>
</tr>
<tr>
<td>▪ More variable, unstable for small data sets</td>
<td>▪ Disadvantages</td>
</tr>
<tr>
<td>▪ Limited to simple models (e.g., linear regression)</td>
<td>▪ Must assume parameteric forms</td>
</tr>
<tr>
<td></td>
<td>▪ Computationally intensive</td>
</tr>
</tbody>
</table>

6/15/11 Brandon C. Kelly, bckelly@cfa.harvard.edu
Some Measurement Error Model References

- **Text books:**

- **Astro articles:**