514.05. Statistical Analysis of Astronomical Measurement Errors

Richard A. Bilonick, PhD

R. A. Bilonick Statistics Consultancy LLC

rabilonick@gmail.com

237th Meeting of the American Astronomical Society

January 15, 2021

514.05

2021-02-01

514.05. Statistical Analysis of Astronomical Measurement Errors

Richard A. Bilonick, PhD

R. A. Bilonick Statistics Consultancy LLC

rabilonick@gmail.com

237th Meeting of the American Astronomical Society

January 15, 2021

Quantifying Agreement

The goal is to decompose the total error into the sum of systematic error (bias) and random error (imprecision).

Bias can be either constant or nonconstant. A scale bias is one kind of nonconstant bias due to different units of measurement.

Random errors can be characterized by a probability distribution and the imprecision gauged by a standard deviation.

Quantifying Agreement
The goal is to decompose the total error into the sum of systematic error (bias) and random error (imprecision).
Bias can be either constant or nonconstant. A scale bias is one kind of nonconstant bias due to different units of measurement.
Random errors can be characterized by a probability distribution and the imprecision gauged by a standard deviation.

1. Using measurement error models is important because many researchers use methods that just plain are wrong.
2. If you get the calibration curve wrong, you will have a false picture of how methods compare.
3. Ordinary regression rarely will produce a curve is even approximately close to the correct calibration curve.
4. And there are TWO different regression lines to boot! They both can't be correct!

DEEP2 Spectroscopic & Photometric Data

Galaxies	1432
z_{spec}	Spectroscopy
z_{fink}	S. Finkelstein
z_{font}	A. Fontana
z_{pfor}	J. Pfor
z_{salv}	M. Salvato
z_{wikl}	T. Wiklind
z_{wuyt}	S. Wuyts

Funding for the DEEP2 Galaxy Redshift Survey has been provided by NSF grants AST-95-09298, AST-0071048, AST-0507428, and AST-0507483 as well as NASA LTSA grant NNG04GC89G. Jeffrey Newman and Ditran Kodra kindly provided the redshift data.

1. There were 1432 galaxies which had their redshift measured by all seven methods.
2. A one-factor measurement error model would not work for all seven methods because the six photometric methods all depend on the same photometric information.
3. Even a two-factor model will not work here because the spectroscopic measurements were not repeated making it impossible to separate the true redshift variance from the random error.

Ordinary Regression Model

In an ordinary regression model, one variable must be a response (dependent) variable and the other must be an explanatory (independent) variable which is measured WITHOUT random error.

The decision to use ordinary regression limits you to two measurement methods at a time.

The decision on dependent/independent variables is consequential if the independent variable is contaminated by random error.

Depending on which of the 2 methods is chosen as the independent variable, you will get a different regression line, i.e., there are ALWAYS 2 regression lines (so long at the correlation does not equal 1 or -1).

It is well known that regressing against an independent variable with random measurement error biases the slope estimate toward zero. The greater the random error, the greater the bias toward zero.

Ordinary Regression Model

In an ordinary regression model, one variable must be a response (dependent) variable and the other must be an explanatory (independent) variable which is measured WITHOUT random error.

The decision to use ordinary regression limits you to two measurement methods at a time.

The decision on dependent/independent variables is consequential if the independent variable is contaminated by random error.

Depending on which of the 2 methods is chosen as the independent variable, you will get a different regression line, i.e., there are ALWAYS 2 regression lines (so long at the correlation does not equal 1 or -1).

It is well known that regressing against an independent variable with random measurement error biases the slope estimate toward zero. The greater the random error, the greater the bias toward zero.

1. At minimum, pairwise regression would be inefficient and would not use all available information.
2. You should not use ordinary regression when both methods are contaminated by random error.
3. All methods are contaminated by random error to one extent or another. It's always best to account for it.

Measurement Error Model

In a measurement error model, all measurement methods are treated as responses (dependent) variables.

They are dependent on one or more latent (unobserved) variables.

The number of latent variables and their relationships with the measurements depends on how the data were collected.

The measurement error model can be implemented using a structural equation model (SEM). Relationships among the latent variables represents the structure.

Measurement Error Model
In a measurement error model, all measurement methods are treated as responses (dependent) variables.
They are dependent on one or more latent (unobserved) variables.
The number of latent variables and their relationships with the measurements depends on how the data were collected.
The measurement error model can be implemented using a structural equation model (SEM). Relationships among the latent variables represents the structure.

1. The latent variable for the true but unknown redshift represents the assumption that all methods are trying to measure the same quantity.
2. The structural equation model is essentially a set of simultaneous equations with distributional assumptions for the random error and the latent variable(s)
3. When determining the calibration equations, the latent variable vanishes.

Simple One-Factor Measurement Error Model – Path Diagram

The path diagram shows the relationship between the true but unknown redshift ζ and the observed measurements Z_1, Z_2, and Z_3.

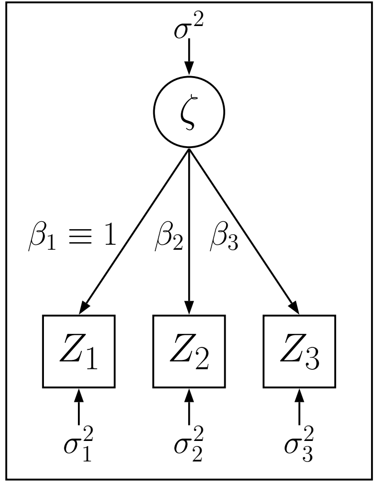
This type of one-factor model would be reasonable for either independent repeats using the same method or three different photometric redshift methods.

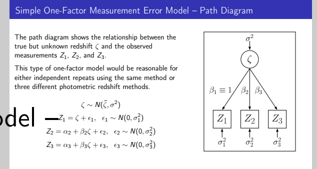
$$\zeta \sim N(\bar{\zeta}, \sigma^2)$$

$$Z_1 = \zeta + \epsilon_1, \quad \epsilon_1 \sim N(0, \sigma_1^2)$$

$$Z_2 = \alpha_2 + \beta_2\zeta + \epsilon_2, \quad \epsilon_2 \sim N(0, \sigma_2^2)$$

$$Z_3 = \alpha_3 + \beta_3\zeta + \epsilon_3, \quad \epsilon_3 \sim N(0, \sigma_3^2)$$





1. Path diagrams are a great way to illustrate the measurement error model or any SEM.
2. However, when the model is very large, it can be difficult to represent the entire model by a path diagram.
3. If you can draw the path diagram it will help guide you in writing the SEM code.

Calibration Equations

Using some simple algebra and taking expectations, the calibration curve between any two types of photometric measurements Z_i and $Z_{i'}$ can be easily derived as:

$$E[Z_i] = \left[\alpha_i - \frac{\beta_i}{\beta_{i'}}\alpha_{i'}\right] + \left(\frac{\beta_i}{\beta_{i'}}\right) E[Z_{i'}]$$

The estimation of the parameters clearly requires a nonlinear methodology.

Calibration Equations

Using some simple algebra and taking expectations, the calibration curve between any two types of photometric measurements Z_i and $Z_{i'}$ can be easily derived as:

$$E[Z_i] = \left[\alpha_i - \frac{\beta_i}{\beta_{i'}}\alpha_{i'}\right] + \left(\frac{\beta_i}{\beta_{i'}}\right) E[Z_{i'}]$$

The estimation of the parameters clearly requires a nonlinear methodology.

1. While there are always two DIFFERENT regreassion equations for any pair of methods, there is only ONE calibration curve relating the methods.
2. You must have at least three measurements for each object in order to estimate all the model parameters.
3. These three or more measurements per object could be three or more different methods or fewer methods if replaced by independent repeats of some of the other methods.

Method of Moments Estimators for 3 Methods

The method of moments equates the observed covariance matrix **S** to the theoretical covariance matrix Σ resulting in six equations (setting $\beta_1 = 1$) in six unknowns:

$$
\begin{aligned}
S_1^2 &= \sigma^2 + \sigma_1^2 \\
S_2^2 &= \beta_2^2\sigma^2 + \sigma_2^2 \\
S_3^2 &= \beta_3^2\sigma^2 + \sigma_3^2 \\
S_{12} &= \beta_2\sigma^2 \\
S_{13} &= \beta_3\sigma^2 \\
S_{23} &= \beta_2\beta_3\sigma^2
\end{aligned}
$$

Method of Moments Estimators for 3 Methods
The method of moments equates the observed covariance matrix S to the theoretical covariance matrix Σ resulting in six equations (setting β1 = 1) in six unknowns:

1. Here we have six eqations in six unknowns for the scale parameters and variance components.
2. Because the number of eqations equals the number of parameters we get a unique solution.
3. In this case, we can not get a gauge of how well the model fits because there will be zero degree of freedom. This is not a big problem but it's always better to have more than three methods an or repeats - the more the better.

Method of Moments

This easily leads to the solution for the method of moments parameter estimates:

$$\beta_2 = S_{23}/S_{13}$$
$$\beta_3 = S_{23}/S_{12}$$
$$\sigma_1^2 = S_1^2 - \frac{S_{12}S_{13}}{S_{23}}$$
$$\sigma_2^2 = S_2^2 - \frac{S_{12}S_{23}}{S_{13}}$$
$$\sigma_3^2 = S_3^2 - \frac{S_{13}S_{23}}{S_{12}}$$
$$\sigma^2 = \frac{S_{12}S_{13}}{S_{23}}$$

Note that β_1 was constrained to 1.

Method of Moments

This easily leads to the solution for the method of moments parameter estimates:

$$\beta_2 = S_{23}/S_{13}$$
$$\beta_3 = S_{23}/S_{12}$$
$$\sigma_1^2 = S_1^2 - \frac{S_{12}S_{13}}{S_{23}}$$
$$\sigma_2^2 = S_2^2 - \frac{S_{12}S_{23}}{S_{13}}$$
$$\sigma_3^2 = S_3^2 - \frac{S_{13}S_{23}}{S_{12}}$$
$$\sigma^2 = \frac{S_{12}S_{13}}{S_{23}}$$

Note that β_1 was constrained to 1.

1. Here the solution is relatively easy.
2. In practice, we would prefer to use maximum likelihood to estimate the parameters.

Method of Moments – Solutions for αs and $\bar{\zeta}$

Setting the sample means ($\bar{Z}_i$) equal to their expected values given the model:

$$\begin{aligned}
\bar{Z}_1 &= \bar{\zeta} \\
\bar{Z}_2 &= \alpha_2 + \beta_2 \bar{\zeta} \\
\bar{Z}_3 &= \alpha_3 + \beta_3 \bar{\zeta}
\end{aligned}$$

which yielded the following solution:

$$\begin{aligned}
\bar{\zeta} &= \bar{Z}_1 \\
\alpha_2 &= \bar{Z}_2 - \frac{S_{23}}{S_{13}} \bar{Z}_1 \\
\alpha_3 &= \bar{Z}_3 - \frac{S_{23}}{S_{12}} \bar{Z}_1
\end{aligned}$$

Note that α_1 was constrained to 0.

1. We solve for the intercepts separately.
2. We need the scale parameter estimates to complete the intercept estimates.
3. We need both the intercepts and the scale parameter estimates to determine the calibration curve

Method of Moments Estimates – Finkelstein (Z_1), Fontana (Z_2), & Pfor (Z_3)

$$\begin{bmatrix} S_1^2 = 0.165 & S_{12} = 0.157 & S_{13} = 0.15 \\ & S_2^2 = 0.185 & S_{23} = 0.154 \\ & & S_3^2 = 0.173 \end{bmatrix}$$

$$\bar{Z}_1 = \bar{\zeta} = 0.7689$$
$$\bar{Z}_2 = \alpha_2 + \beta_2\bar{\zeta} = 0.7852$$
$$\bar{Z}_3 = \alpha_3 + \beta_3\bar{\zeta} = 0.7849$$

$$\beta_2 = S_{23}/S_{13} = 1.029$$
$$\beta_3 = S_{23}/S_{12} = 0.983$$
$$\sigma_1^2 = S_1^2 - \frac{S_{12}S_{13}}{S_{23}} = 0.01196, \ \sigma_1 = 0.109$$
$$\sigma_2^2 = S_2^2 - \frac{S_{12}S_{23}}{S_{13}} = 0.023, \ \sigma_2/\beta_2 = 0.148$$
$$\sigma_3^2 = S_3^2 - \frac{S_{13}S_{23}}{S_{12}} = 0.02542, \ \sigma_3/\beta_3 = 0.162$$
$$\sigma^2 = \frac{S_{12}S_{13}}{S_{23}} = 0.153, \ \sigma = 0.391$$

$$\bar{\zeta} = \bar{Z}_1 = 0.769$$
$$\alpha_2 = \bar{Z}_2 - \frac{S_{23}}{S_{13}}\bar{Z}_1 = -0.006$$
$$\alpha_3 = \bar{Z}_3 - \frac{S_{23}}{S_{12}}\bar{Z}_1 = 0.029$$

1. The solutions are easy enough to do by hand but usually we would use a computer program.

Results Using `merror` R Package – Converted to $\beta_1 = 1$ Constraint

	n	σ_i	σ_{σ_i}	α_i	β_i	LB	UB	$\frac{\sigma_i}{\beta_i}$
Z_{fink}	1424	0.109	0.031	-0.008	0.996	0.101	0.118	0.110
Z_{font}	1424	0.152	0.035	-0.014	1.025	0.145	0.161	0.149
Z_{pfor}	1424	0.159	0.036	0.021	0.979	0.152	0.168	0.163
True Redshift	1424	0.392	0.077			0.378	0.408	

The `merror` package uses the constraint that the geometric average of the βs equals 1. To convert this so that instead, $\beta_1 = 1$, we divide each β by $\beta_1 = 0.996$:

Method	σ_i	$\beta_1\beta_2\beta_3 = 1$	$\beta_1 = 1$	σ_i/β_i
Finkelstein	0.109	0.996	1.000	0.109
Fontana	0.152	1.025	1.029	0.148
Pfor	0.159	0.979	0.983	0.162

Results Using `merror` R Package – Converted to $\beta_1 = 1$ Constraint

1. There are many ways to impose the constraints.
2. Which ever way you impose the constraints, the conclusions in relative terms are always the same
3. The calibration curves are invariant to the constraints.

Results Using OpenMx R Package

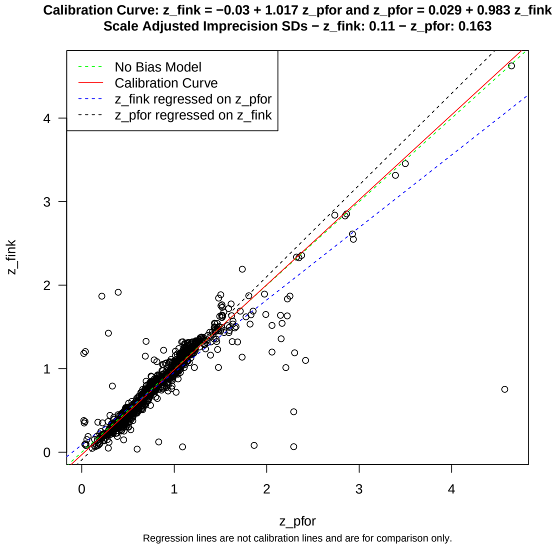
Parameter estimates for a single-factor model for only photometric redshifts for Finkelstein, Fontana, and Pfor. These estimates are the same as previously estimated by the method of moments. Bootstrap sampling (bias corrected) was used to approximate the 95% intervals.

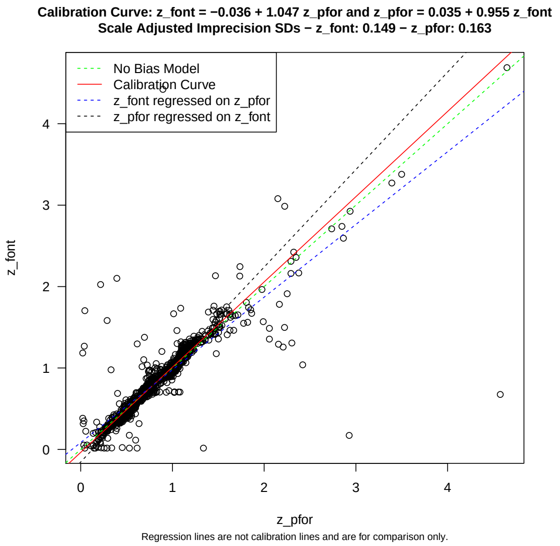
	Lower	Estimate	Upper
		Scale Bias	
β_{fink}		1.000	
β_{font}	0.980	1.029	1.090
β_{pfor}	0.941	0.983	1.021
		Bias Intercept	
α_{fink}		0.000	
α_{font}	-0.049	-0.006	0.028
α_{pfor}	0.004	0.029	0.057

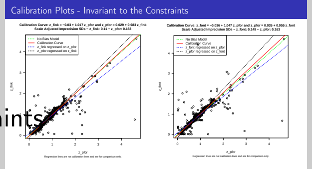
	Lower	Estimate	Upper
		SD for True Redshift Values	
σ_p	0.359	0.391	0.416
		Imprecision SDs	
σ_{fink}	0.064	0.109	0.132
$\sigma_{font}/\beta_{font}$	0.082	0.148	0.185
$\sigma_{pfor}/\beta_{pfor}$	0.091	0.162	0.192

1. The OpenMx model yields the same results as the merror package – so long as the same constraints are used in both.
2. Bootstrapping should provide more realistic confidence intervals.
3. OpenMx can provide confidence intervals for any function of the parameters.

Calibration Plots - Invariant to the Constraints







1. The calibration curves show how wrong ordinary regression can be.
2. You won't know exactly how wrong the ordinary regression curves will be until you fit the correct measurement error model.
3. In my experience, researchers often use the regression curve the inferior of the two curves.

Conclusions

- Ordinary regression causes the estimate of the scale bias to be biased toward 0. This bias was large for the photometric methods. (It also occurs with spectroscopic redshift, but is smaller in magnitude).
- Before you can compare method imprecision, the imprecision standard deviations must be divided by their corresponding scale biases (βs).
- At minimum, you need 3 methods compared on the same objects, or if only two methods, at least one must have independent repeats. (Otherwise, you would be forced to make unverifiable assumptions in order to reduce the number of parameters needed to be estimated.)
- Although 3 methods is the minimum, the more methods compared on the same objects the better.
- It is not necessary to have independent repeats so long as you have at least 3 different methods, but having repeats can be useful.
- A one-factor model is not reasonable if you want to include both spectroscopic and photometric methods in the same model. (See 235.06 for an example of including both spectroscopic and photometric methods in the same comparison - more than one latent factor is needed.)

└Conclusions

Conclusions

- Ordinary regression causes the estimate of the scale bias to be biased toward 0. This bias was large for the photometric methods. (It also occurs with spectroscopic redshift, but is smaller in magnitude).
- Before you can compare method imprecision, the imprecision standard deviations must be divided by their corresponding scale biases (βs).
- At minimum, you need 3 methods compared on the same objects, or if only two methods, at least one must have independent repeats. (Otherwise, you would be forced to make unverifiable assumptions in order to reduce the number of parameters needed to be estimated.)
- Although 3 methods is the minimum, the more methods compared on the same objects the better.
- It is not necessary to have independent repeats so long as you have at least 3 different methods, but having repeats can be useful.
- A one-factor model is not reasonable if you want to include both spectroscopic and photometric methods in the same model. (See 235.06 for an example of including both spectroscopic and photometric methods in the same comparison - more than one latent factor is needed.)

References I

- Richard A. Bilonick, *mirror: Accuracy and Precision of Measurements*, 2015, R package version 2.0.2.
- Steven M. Boker, Michael C. Neale, Hermine H. Maes, Michael J. Wilde, Michael Spiegel, Timothy R. Brick, Ryne Estabrook, Timothy C. Bates, Paras Mehta, Timo von Oertzen, Ross J. Gore, Michael D. Hunter, Daniel C. Hackett, Julian Karch, Andreas M. Brandmaier, Joshua N. Pritikin, Mahsa Zahery, Robert M. Kirkpatrick, Yang Wang, Charles Driver, Massachusetts Institute of Technology, S. G. Johnson, Association for Computing Machinery, Dieter Kraft, Stefan Wilhelm, Sarah Medland, Carl F. Falk, Matt Keller, Manjunath B G, and The Regents of the University of California., *Openmx 2.9.9 user guide*, 2018.
- Michael D. Hunter, *State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment*, Structural Equation Modeling (in press), 1–18.

References I

Richard A. Bilonick, *mirror: Accuracy and Precision of Measurements*, 2015, R package version 2.0.2.

Steven M. Boker, Michael C. Neale, Hermine H. Maes, Michael J. Wilde, Michael Spiegel, Timothy R. Brick, Ryne Estabrook, Timothy C. Bates, Paras Mehta, Timo von Oertzen, Ross J. Gore, Michael D. Hunter, Daniel C. Hackett, Julian Karch, Andreas M. Brandmaier, Joshua N. Pritikin, Mahsa Zahery, Robert M. Kirkpatrick, Yang Wang, Charles Driver, Massachusetts Institute of Technology, S. G. Johnson, Association for Computing Machinery, Dieter Kraft, Stefan Wilhelm, Sarah Medland, Carl F. Falk, Matt Keller, Manjunath B G, and The Regents of the University of California., *Openmx 2.9.9 user guide*, 2018.

Michael D. Hunter, *State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment*, Structural Equation Modeling (in press), 1–18.

- JL Jaech, *Statistical Analysis of Measurement Errors, John Wiley & Sons*, Inc., New York, NY, USA (1985).
- Michael C. Neale, Michael D. Hunter, Joshua N. Pritikin, Mahsa Zahery, Timothy R. Brick, Robert M. Kirkpatrick, Ryne Estabrook, Timothy C. Bates, Hermine H. Maes, and Steven M. Boker, *OpenMx 2.0: Extended structural equation and statistical modeling*, Psychometrika **81** (2016), no. 2, 535–549.
- Joshua N. Pritikin, Michael D. Hunter, and Steven M. Boker, *Modular Open-source Software for Item Factor Analysis*, Educational and Psychological Measurement **75** (2015), no. 3, 458–474.
- R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018.

References II

JL Jaech, *Statistical Analysis of Measurement Errors, John Wiley & Sons*, Inc., New York, NY, USA (1985).
Michael C. Neale, Michael D. Hunter, Joshua N. Pritikin, Mahsa Zahery, Timothy R. Brick, Robert M. Kirkpatrick, Ryne Estabrook, Timothy C. Bates, Hermine H. Maes, and Steven M. Boker, *OpenMx 2.0: Extended structural equation and statistical modeling*, Psychometrika **81** (2016), no. 2, 535–549.
Joshua N. Pritikin, Michael D. Hunter, and Steven M. Boker, *Modular Open-source Software for Item Factor Analysis*, Educational and Psychological Measurement **75** (2015), no. 3, 458–474.
R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018.