Commentary

Chad M. Schafer

Carnegie Mellon University, Department of Statistics & Data Science

cschafer@cmu.edu

Systematic Uncertainty as Model Misspecification

- Two classes of systematic uncertainty:

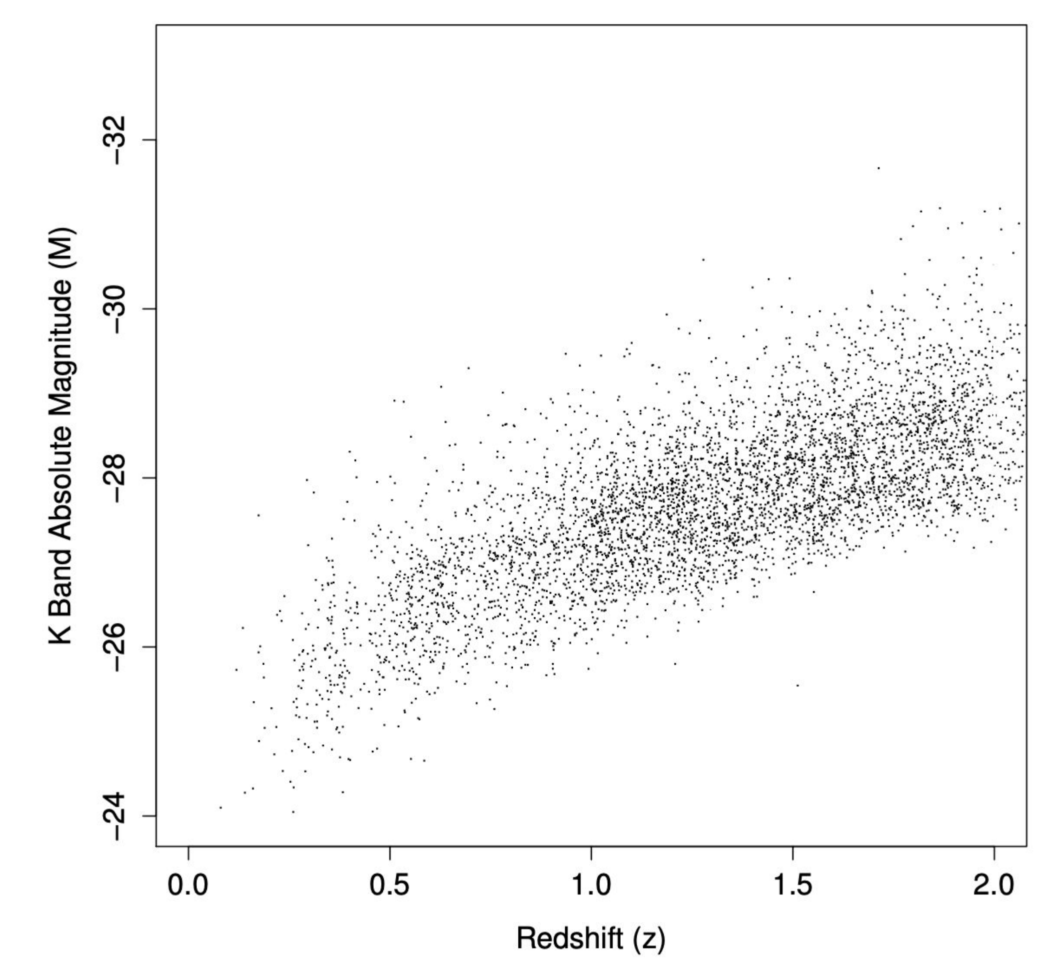
  1. Uncertainties due to analysis choices, which are typically binary variations (chosen as an example of other possible options).

     If there are multiple choices of model for the signal, *e.g.* choice of resonances in an amplitude fit, it is usual to quote fit results for every model that has a similar fit quality.

  2. Statistical uncertainties on external inputs.

     e.g. due to limited MC sample sizes or calibration sample sizes.

Cosmological Context



Model the bivariate distribution

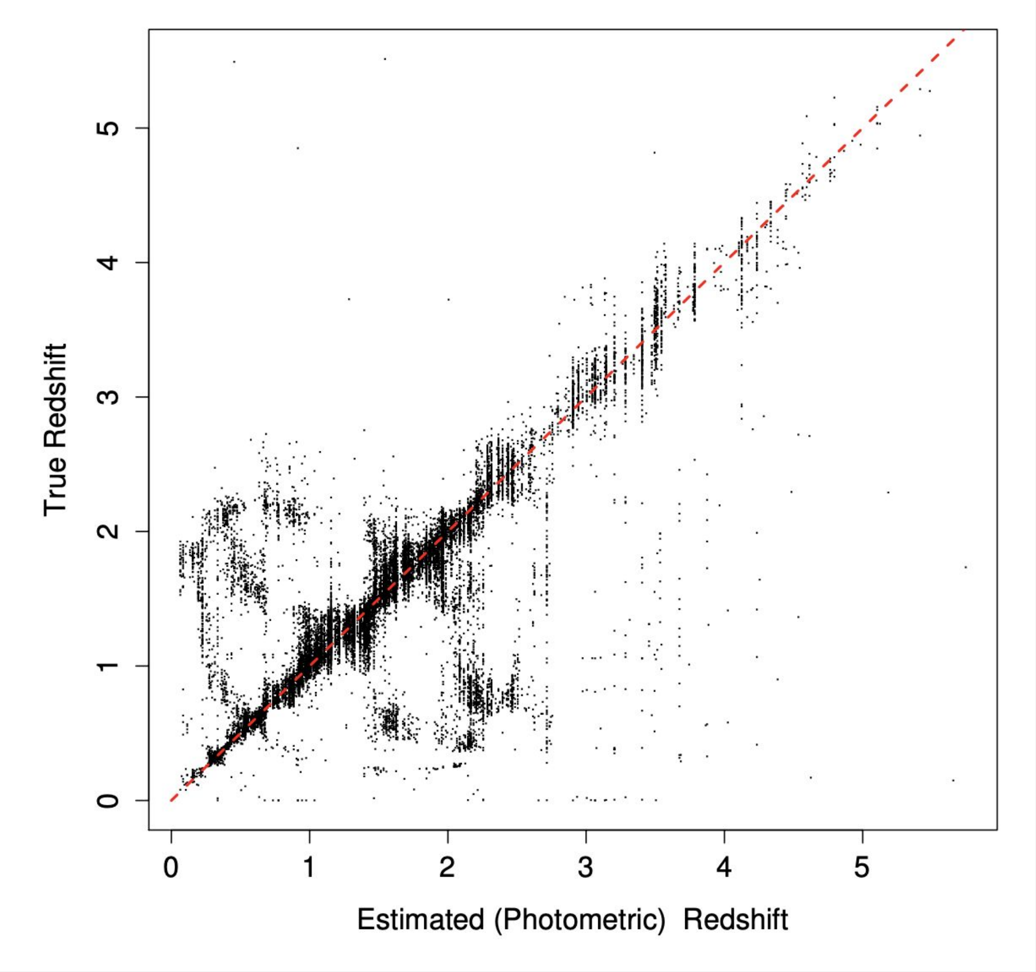
(5000 quasars taken from Peth (2010) sample)

Cosmological Context

Year	Count	Source
1984	35	Marshall, et al.
1988	420	Boyle, et al.
1995	1,200	Pei
2002	3,814	Schneider, et al. (SDSS Early Data Release)
2003	16,713	Schneider, et al. (SDSS DR1)
2005	46,420	Schneider, et al. (SDSS DR3)
2007	77,429	Schneider, et al. (SDSS DR5)
2010	105,783	Schneider, et al. (SDSS DR7)
2018	526,356	Paris, et al. (SDSS DR14)
2017	2.0 million	MILLIQUAS, Version 5.2
2030	> 10 million	LSST

Progression of quasar sample sizes

Cosmological Context



Model the bivariate distribution as a function of key cosmological parameters

(5000 quasars taken from Peth (2010) sample)

Cosmological Context

Modern cosmology motivates use of...

...Bayesian methods

...Nonparametric methods

...likelihood-free methods

Cosmological Context

Modern cosmology motivates use of...

...Bayesian methods

...Nonparametric methods

...likelihood-free methods

What is missing?

Issues surrounding discovery claims

Discovery Issues

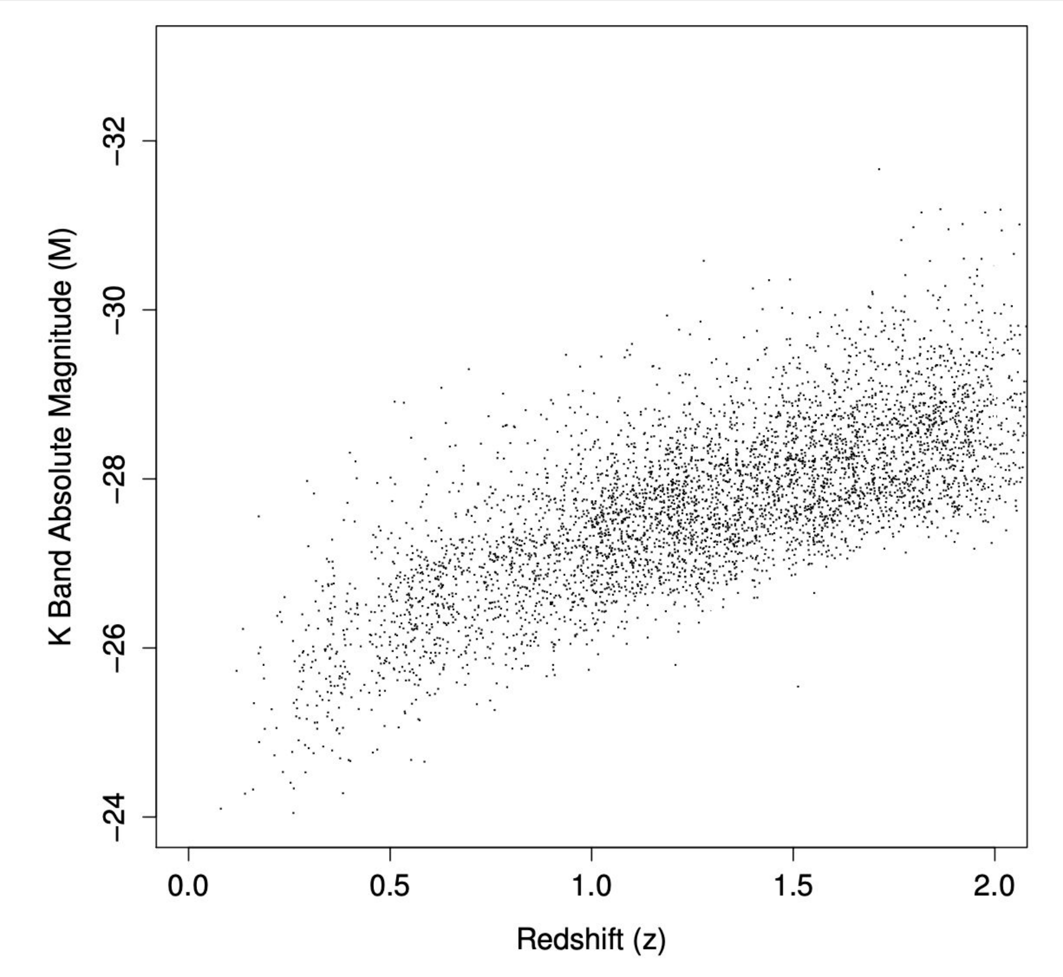
Does the need for agreement on when a discovery has been made unduly constrain the space of allowable analysis methods?

Does the focus on preserving power (sensitivity) bias significance results?

Cowan, et al. (2011)

“For the significances obtained to be valid it is necessary that the model predictions for data distributions represent accurately the underlying theory being tested. That is, any errors due to approximations, e.g., in detector modeling or in methods used to relate observable quantities to the fundamental theories, should be negligible for some point in the full parameter space. **By including additional parameters in the model one can approach this ideal situation more closely. However, the additional flexibility introduced to parameterize systematic effects results, as it should, in a loss in sensitivity.**” (page 2)

Back to cosmology



Model the bivariate distribution

(5000 quasars taken from Peth (2010) sample)

Handling Model Misspecification

Handling uncertainties in background shapes: the discrete profiling method

P. D. Dauncey[a]*, M. Kenzie[b], N. Wardle[b] and G. J. Davies[a]

[a] *Department of Physics, Imperial College London, Prince Consort Road, London, SW7 2AZ, UK.*
[b] *CERN, CH-1211 Geneva 23, Switzerland.*
E-mail: `P.Dauncey@imperial.ac.uk`

ABSTRACT: A common problem in data analysis is that the functional form, as well as the parameter values, of the underlying model which should describe a dataset is not known *a priori*. In these cases some extra uncertainty must be assigned to the extracted parameters of interest due to lack of exact knowledge of the functional form of the model. A method for assigning an appropriate error is presented. The method is based on considering the choice of functional form as a discrete nuisance parameter which is profiled in an analogous way to continuous nuisance parameters. The bias and coverage of this method are shown to be good when applied to a realistic example.