Peelle's Pertinent Puzzle and D'Agostini Bias — *Estimating the Mean with Relative Systematic Uncertainty*

Speaker: Mike Grosskopf

Co-authors:

Scott Vander Wiel

Denise Neudecker

Peelle's Pertinent Puzzle (PPP)

- In 1987 at Oak Ridge National Lab, R.W. Peelle described having two observations to estimate a shared mean:

$$y_1 = 1.0 \pm 10\% \qquad y_2 = 1.5 \pm 10\%$$

and 20% fully correlated uncertainty

- Uncertainties
  - "Fully correlated" means multiplicatively up or down together.
  - Three uncertainties are taken to be independent.
  - Relative uncertainties are assumed to be "1 sigma" values
- Peelle was a physicist doing nuclear data evaluation and he worked out the generalized least squares (GLS) estimate of the mean, as is standard in the field

Peelle's Pertinent Puzzle (PPP)

- Generalized least squares as per Peelle:

$$\boldsymbol{y} = (1.0, 1.5)^T$$

$$\Sigma = 0.1^2 \cdot \mathrm{diag}(\boldsymbol{y}^2) \; + \; 0.2^2 \cdot \boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$$

$$\hat{\mu} = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \boldsymbol{y}$$

$$\hat{\sigma}^2 = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1}$$

- And the answer is …

Los Alamos NATIONAL LABORATORY

Peelle's Pertinent Puzzle (PPP)

- Generalized least squares as per Peelle:

$$\boldsymbol{y} = (1.0, 1.5)^T$$

$$\Sigma = 0.1^2 \cdot \mathrm{diag}(\boldsymbol{y}^2) \; + \; 0.2^2 \cdot \boldsymbol{y}\boldsymbol{y}^\mathrm{T}$$

$$\hat{\mu} = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \boldsymbol{y}$$

$$\hat{\sigma}^2 = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1}$$

- And the answer is …

$$\hat{\mu} = 0.88 \pm 0.25$$

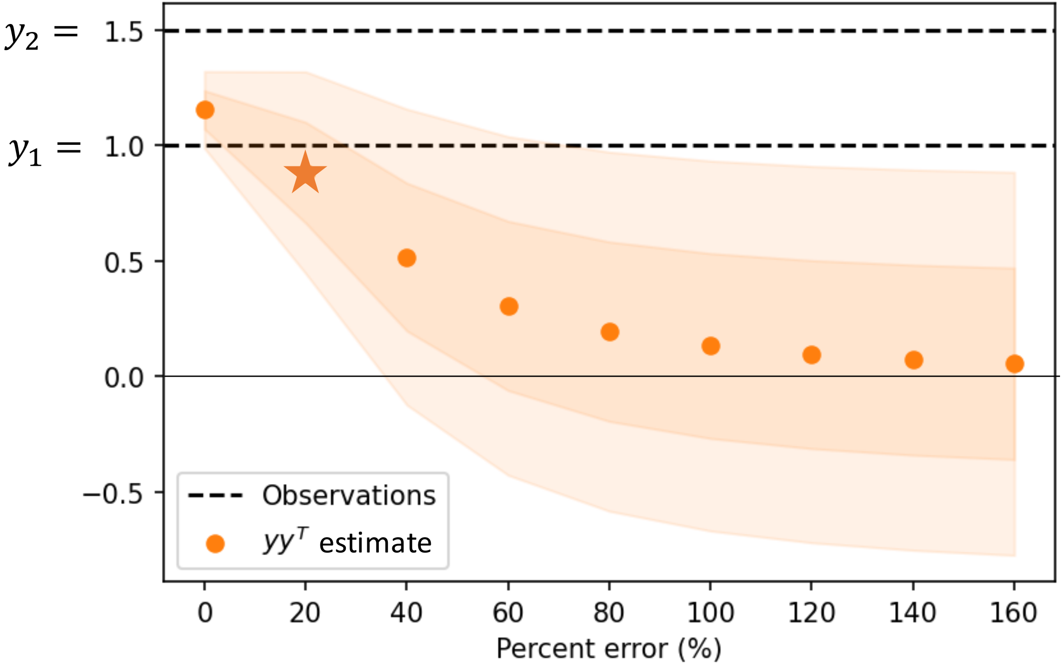
below **both** observations!

- Terminology: *Data Weighted – Generalized Least Squares (**DW-GLS**)* refers to GLS with $\boldsymbol{y}$ in the assumed covariance matrix
  - Many scientists don't consider having $\boldsymbol{y}$ in its own covariance as odd

Los Alamos NATIONAL LABORATORY

As fully-correlated error increases we expect

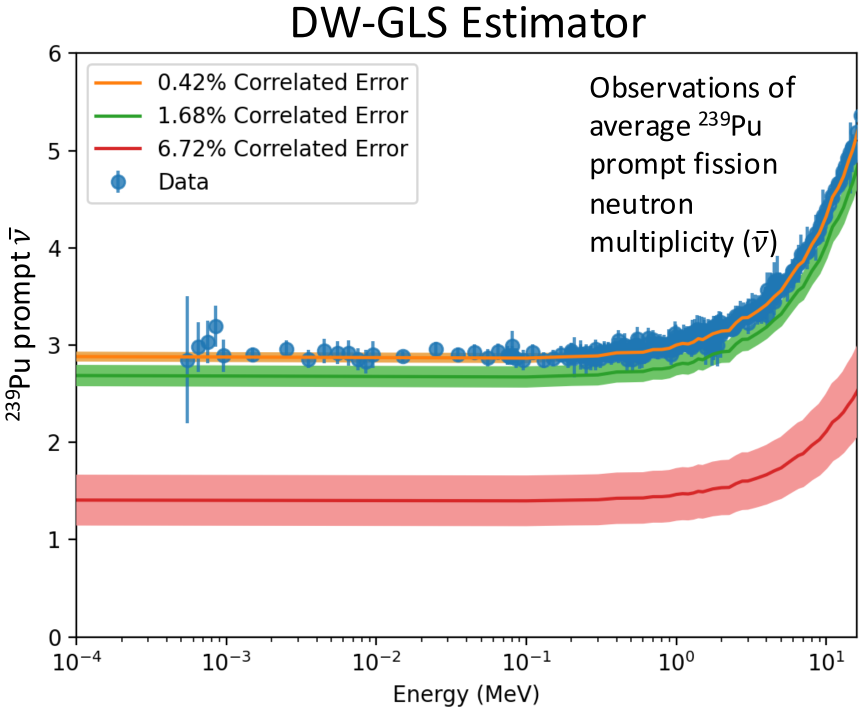
- *proportional* increase in uncertainty,
- *and not* additional bias…



… but neither holds with DW-GLS

Los Alamos NATIONAL LABORATORY

This problem exists more widely when using generalized least squares for nuclear data evaluation



Fitted curves (prior not shown)

- **Nominal case** has 0.42% fully-correlated error and looks OK but
  - fitted is 0.7% below data
  - 73% of residuals are positive
- Increasing the fully-correlated error by **4x** and **16x** reveals dramatic bias

Los Alamos NATIONAL LABORATORY

(Data from Marini et al., Phys Letters B, 835, 2022,137513)

High-Energy Physicists know this as *D'Agostini Bias*

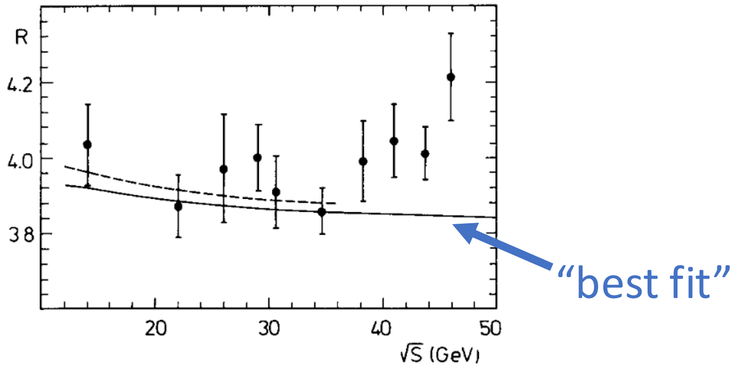


Fig. 2. R measurements from PETRA and PEP experiments with the best fits of QED + QCD to all the data (full line) and only below 36 GeV (dashed line). All data points are correlated (see text).

Example of a recent non-explanation

“The best-fit undershoots the data, essentially because with multiplicative uncertainties a lower prediction has a smaller uncertainty [D’Agostini, 1994].”

D'Agostini, Giulio. "On the use of the covariance matrix to fit correlated data." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 346.1-2 (**1994**): 306-311.

Ball, Richard D., et al. "Precision determination of the strong coupling constant within a global PDF analysis: NNPDF Collaboration." *The European Physical Journal* C 78 (**2018**): 1-16.

What is going on in PPP?

- **Is this not a problem at all?**
  - Two observations with correlated errors aren't at all unlikely to fall on the same side of the mean
    - Why biased low rather than high? Symmetry would seem to make a case for no preference in either direction
    - Why so far outside the data? **The estimate makes one observation a 3σ outlier**

- Is this a problem with the GLS estimator for relative, correlated errors?

- Is this a fluke of the two-observation problem?

Los Alamos NATIONAL LABORATORY

What is going on here?

- Is this not a problem at all?
  - Two observations with correlated errors aren't at all unlikely to fall on the same side of the mean
- **Is this a problem with the GLS estimator for relative, correlated errors?**
  - Let's explore a couple other estimators:
    - Maximum likelihood estimator: $\boldsymbol{y} \sim \mathrm{MVN}(\boldsymbol{\mu}, \mathrm{diag}(0.1^2\boldsymbol{\mu}^2) + 0.2^2\boldsymbol{\mu}\boldsymbol{\mu}^\mathrm{T})$
    - Iteratively re-weighted least squares: Do GLS, but iterate from initial $\hat{\boldsymbol{\mu}} = \boldsymbol{y}$ using $\Sigma = \mathrm{diag}(0.1^2\, \hat{\boldsymbol{\mu}}^2) + 0.2^2\, \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\mathrm{T}$
- Is this a fluke of the two-observation problem?

Los Alamos NATIONAL LABORATORY

What is going on here?

Recall
$y_1 = 1.0 \pm 10\%$
$y_2 = 1.5 \pm 10\%$

- Is this not a problem at all?
  - Two observations with correlated errors aren't at all unlikely to fall on the same side of the mean
- **Is this a problem with the GLS estimator for relative, correlated errors?**
  - Let's explore a couple other estimators:
    - Maximum likelihood estimator: $\hat{\mu}_{MLE} = 1.531$
    - Iteratively re-weighted least squares: $\hat{\mu}_{IRLS} = 1.25$
- Is this a fluke of the two-observation problem?

MLE is **above** both observations!

Observations just touch at $\pm 2\sigma$

MLE links the mean to this spread.

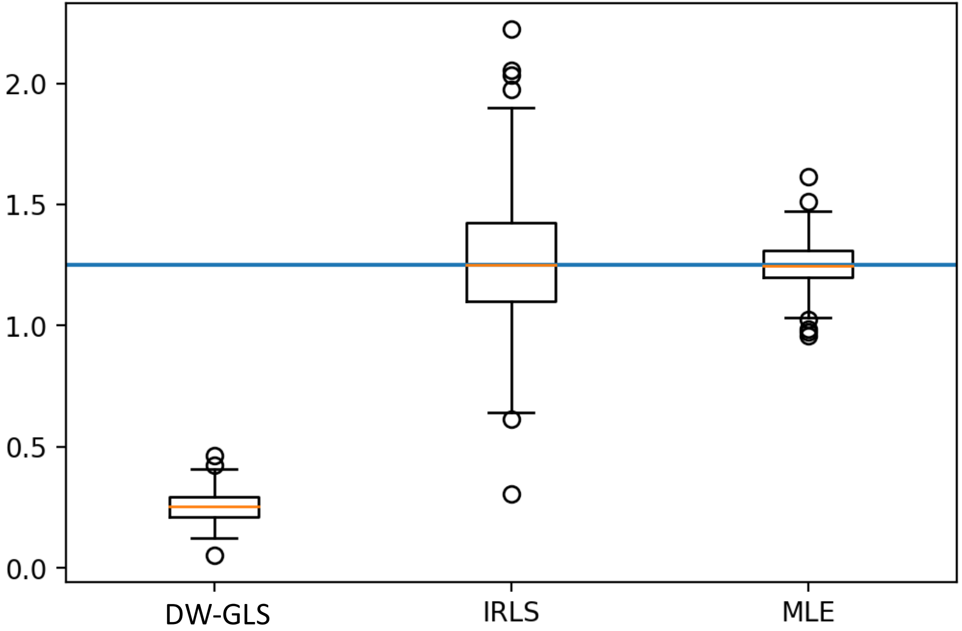
Los Alamos NATIONAL LABORATORY

What is going on here?

- Is this not a problem at all?
  - Two observations with correlated errors aren't at all unlikely to fall on the same side of the mean
- Is this a problem with the GLS estimator for relative, correlated errors?
- **Is this a fluke of the two-observation problem?**
  - Simulate 500 samples, each with
    - *n*=100 observations
    - mean=1.25
    - SDs: 10% independent + 20% fully-correlated

Los Alamos NATIONAL LABORATORY

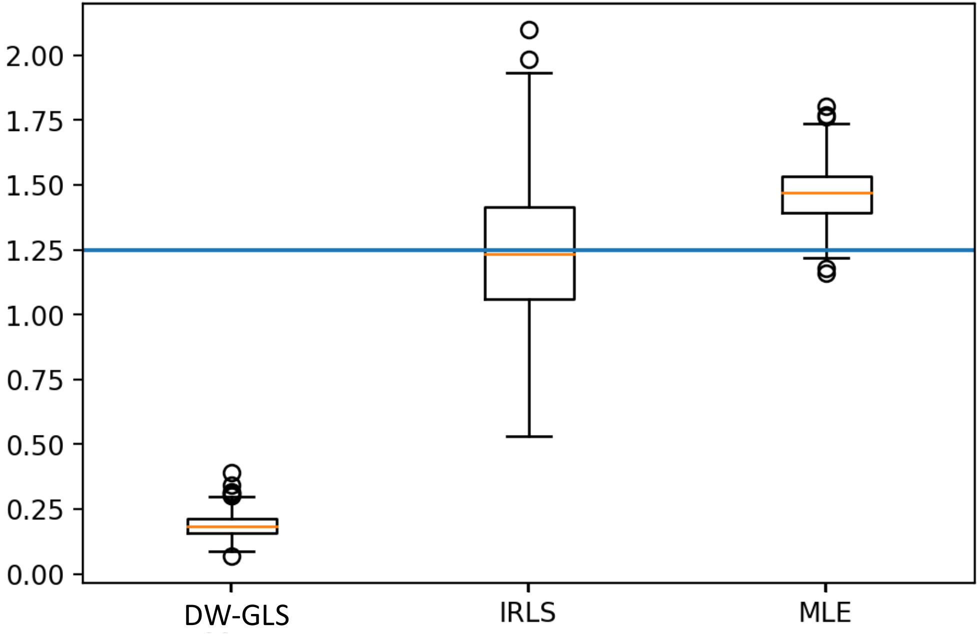
Estimates of the mean parameter from 500 simulated samples



Now the MLE looks great!

Why does the MLE look great now?

Estimates of the mean parameter with data generated with 12% independent error when 10% is assumed



The MLE no longer looks so great

The MLE is sensitive to mis-specified uncertainty!

Los Alamos NATIONAL LABORATORY

Conclusions so far:

- DW-GLS (i.e., using $\sigma^2 \boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$ in the covariance) biases estimates toward zero
  - large $\sigma \Longrightarrow$ bad bias.

- The MLE (and Bayes) is efficient but sensitive to mis-specified uncertainties

- IRLS is unbiased and robust to mis-specified uncertainties

PPP / D'Agostini Bias features in many papers over 30 years

In fact, IRLS has been reinvented to fix the bias

- but it is not recognized as fully legit

E.g., from Capote, *et al.* (2009)

An empirical "fix" compensates for PPP in a practical way as suggested by Chiba and Smith (1991).

References

Badikov, S., Zhenpeng, C., Carlson, A., Gai, G., Hambsch, F., Hofmann, H., Kawano, T., Larson, N., Pronyaev, V., Smith, D. et al. (2007), International evaluation of neutron cross-section standards. iaea report, Technical report, STI/PUB/1291.

Ball, R. D., Del Debbio, L., Forte, S., Guffanti, A., Latorre, J. I., Rojo, J. & Ubiali, M. (2010), 'Fitting parton distribution data with multiplicative normalization uncertainties', *Journal of High Energy Physics* **2010**(5), 1–34.

Benayoun, M., David, P., DelBuono, L. & Jegerlehner, F. (2015), 'Muon $g-2$ estimates: can one trust effective lagrangians and global fits?', *The European Physical Journal C* **75**(12), 1–29.

Burr, T., Graves, T., Hengartner, N., Kawano, T., Pan, F. & Talou, P. (2011), 'Alternatives to the least squares solution to Peelle's Pertinent Puzzle', *Algorithms* **4**(2), 115–130.

Burr, T., Kawano, T., Talou, P., Pan, F. & Hengartner, N. (2011), 'Defense of the least squares solution to Peelle's Pertinent Puzzle', *Algorithms* **4**(1), 28–39.

Capote, R., Badikov, S., Carlson, A., Duran, I., Gunsing, F., Neudecker, D., Pronyaev, V., Schillebeeckx, P., Schnabel, G., Smith, D. & Wallner, A. (2020), 'Unrecognized sources of uncertainties (usu) in experimental nuclear data', *Nuclear Data Sheets* **163**, 191–227.

Carlson, A., Pronyaev, V., Smith, D., Larson, N. M., Chen, Z., Hale, G., Hambsch, F.-J., Gai, E., Oh, S.-Y., Badikov, S. et al. (2009), 'International evaluation of neutron cross section standards', *Nuclear Data Sheets* **110**(12), 3215–3324.

Chiba, S. & Smith, D. L. (1991), A suggested procedure for resolving an anomaly in least-squares data analysis known as "Peelle's Pertinent Puzzle" and the general implications for nuclear data evaluation, Technical report, Argonne National Lab., IL (United States).

D'Agostini, G. (1994), 'On the use of the covariance matrix to fit correlated data', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **346**(1-2), 306–311.

Eskola, K. J., Helenius, I., Kuha, M. & Paukkunen, H. (2020), 'Shadowing in inelastic nucleon-nucleon cross section?', *Physical Review Letters* **125**(21), 212301.

Fruehwirth, R., Neudecker, D. & Leeb, H. (2012), Peelle's pertinent puzzle and its solution, *in* 'EPJ Web of Conferences', Vol. 27, EDP Sciences, p. 00008.

Hanson, K. M., Kawano, T. & Talou, P. (2005), Probabilistic interpretation of Peelle's Pertinent Puzzle and its resolution, *in* 'AIP Conference Proceedings', Vol. 769, American Institute of Physics, pp. 304–307.

Hoid, B.-L., Hoferichter, M. & Kubis, B. (2020), 'Hadronic vacuum polarization and vector-meson resonance parameters $e^+e^- \to \pi^0\gamma$', *The European Physical Journal C* **80**(10), 1–9.

Neudecker, D., Frühwirth, R. & Leeb, H. (2012), 'Peelle's Pertinent Puzzle: A fake due to improper analysis', *Nuclear science and engineering* **170**(1), 54–60.

Oh, S. & Seo, C. (2004), 'Box-Cox transformation for resolving Peelle's Pertinent Puzzle in curve fitting', *PHYSOR-2004* pp. 25–29.

Peelle, R. (1987), 'Peelle's Pertinent Puzzle', *informal ORNL memorandum* **13**.

Pisanti, O. (2020), Improved nuclear reaction network for a reliable estimate of primordial deuterium yield, *in* 'Journal of Physics: Conference Series', Vol. 1468, IOP Publishing, p. 012010.

Smith, D., Badikov, S., Gai, E., Soo-Youl, O., Kawano, T., Larson, N. & Pronyaev, V. (2007), Perspectives on Peelle's Pertinent Puzzle, *in* 'International evaluation of neutron cross-section standards', IAEA.

Why does DW-GLS behave so badly?

Insight from a regression problem

Regress $\mathbf{y}$ on two predictors

$$\mathbf{y} = \alpha\mathbf{1} + \beta\boldsymbol{x} + \gamma\boldsymbol{z} + \boldsymbol{\epsilon}$$

Suppose we care about α, β, but γ is just a nuisance with priors

$$\gamma \sim \mathrm{N}\left(0, \sigma_\gamma^2\right), \qquad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$$

Integrating over γ:

$$\mathbf{y} = \alpha\mathbf{1} + \beta\boldsymbol{x} + \boldsymbol{e}$$
$$\boldsymbol{e} \sim \mathrm{N}\left(\mathbf{0},\ \sigma_\epsilon^2 \boldsymbol{I} + \sigma_\gamma^2 \boldsymbol{z}\boldsymbol{z}^{\boldsymbol{T}}\right)$$

I.e., one-predictor but correlated errors

Application to DW-GLS

DW-GLS uses

$$\Sigma = 0.1^2 \cdot \mathrm{diag}(\boldsymbol{y}^2) + 0.2^2 \cdot \boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$$

and the $\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$ is like regressing $\boldsymbol{y}$ on itself!

With $\boldsymbol{z} = \boldsymbol{y}$, the covariance absorbs the signal and

$$\left(\hat{\alpha}\ \hat{\beta}\right) \to (0,0) \quad \text{as} \quad \sigma_\lambda \to \infty$$

Larger correlated error ⇒ more over-fitting

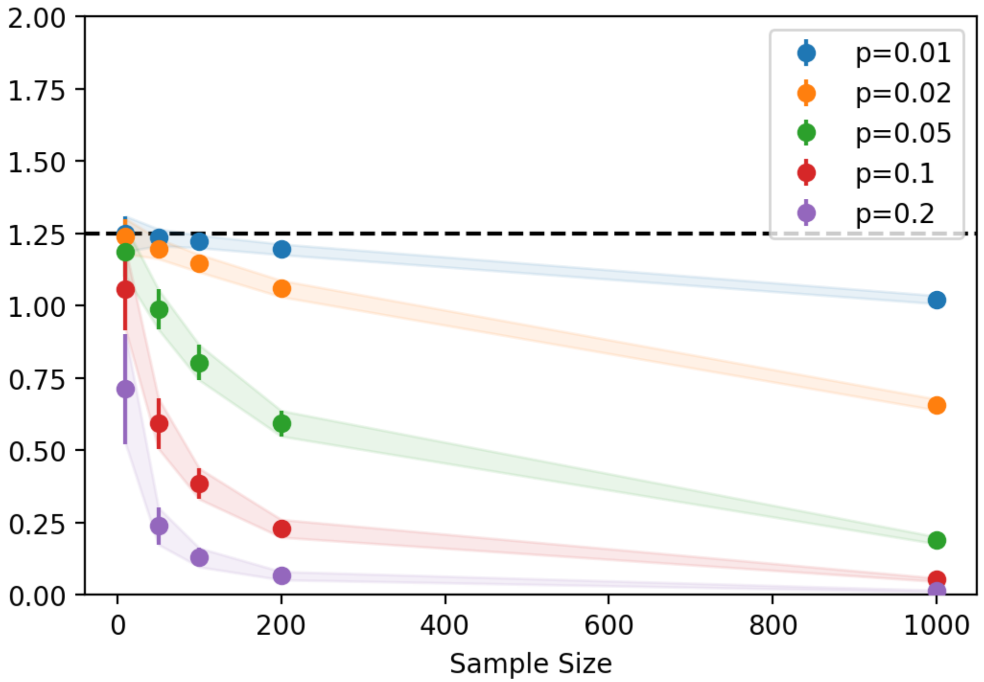
That is, DW-GLS with a fully correlated term $\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$ is inherently biased toward zero.

Los Alamos NATIONAL LABORATORY

Coming back to DW-GLS, is it at least OK for small correlated uncertainty?

No

- It is not consistent.
- Adding more data drives the estimate to 0 for *any* relative correlated uncertainty, *p>0*



Conclusions

Advice for analysts

- **DO NOT** represent fully-correlated relative uncertainty as $p^2 \boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}$
  - equivalent to adding $\boldsymbol{y}$ as a predictor for $\boldsymbol{y}$ itself
  - this drives the signal of interest to zero
  - Results in bad mean AND bad uncertainty
- **DO** represent the uncertainty as $p^2\, \widehat{\boldsymbol{\mu}}\, \widehat{\boldsymbol{\mu}}^{\boldsymbol{T}}$ and use IRLS
  - this is robust to mis-specified p
  - also robust to mis-specified independent error
- **DO** carefully consider how relative error can allow noise to influence the mean estimates with MLE and Bayesian inference

Acknowledgments

- Research reported in this publication was supported by the U.S. Department of Energy LDRD program at Los Alamos National Laboratory.
- NCERC is supported by the DOE Nuclear Criticality Safety Program, funded and managed by the National Nuclear Security Administration for the Department of Energy.
- Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the US Department of Energy under Contract No. 89233218CNA000001.

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

NCSP NUCLEAR CRITICALITY SAFETY PROGRAM

NCERC NATIONAL CRITICALITY EXPERIMENTS RESEARCH CENTER

MCNP6

Los Alamos NATIONAL LABORATORY