Bias-Variance Trade-off In Proton Radius Extractions From Electron Scattering Data
or
How the Bias Bias Biases 'Unbiased' Proton Radius Extractions
or
How I Learned to Stop Worrying and Love the Bias

Randall Evan McClellan, David Meekins, Douglas Weadon Higinbotham, and Christian Weiss
Jefferson Lab, Newport News, VA 23606

William and Mary Group
William and Mary, Williamsburg, VA 23601

Intuitively, it is often assumed that a truer model will necessarily yield a more accurate fit. Herein we revisit this ansatz in the context of extracting the proton radius from idealized charge form factor data. This is done with a simple Monte Carlo simulation following exactly the conditions as outlined in the classic paper that rejected linear model extractions in favor of quadratic extractions on the basis that the simulations clearly indicated bias. In statistical modeling, one not only needs to consider bias but also variance. Once both bias and variance are taken into account via calculation of the root mean square error, one now finds that at low Q^2 the linear model extractions in fact have better predictive power due to significantly lower variance than quadratic models.

Finally we apply the rational function to the entire Mainz spectrometer B data set and get $\chi^2 = 679$ with 681 points fitted for a $\chi^2/\nu = 1.00$.

We then repeat using the rebinned data of Arrington/Hill/Lee, which gives the same answer though suggests that they have indeed now overestimated the uncertainties:

$$\chi^2 = 229.69 \text{ for 333 data points, } \chi^2/\nu = 0.70$$

Pade Output Proton Charge Radius 0.8495479660202573
Pade Output Proton Magnetic Radius 0.8575784681411048
Fitted 333 data points:
180 MeV: Total Chi2 34.179 Total 61.000 chi2/v 0.579

For a proton charge radius of 0.850 ± 0.001 fm using two completely different binnings of the Mainz data. and a result that mimics that shape of asym. data. (try that trick with polynominals ;-)

I. INTRODUCTION

The proton's charge radius, r_p, can be extracted by determining the slope of the electric form factor as Q^2 approaches zero. In fact the relation between this slope and the radius is defined to be

$$G_E(Q^2) = 1 + \sum_{n\geq 1} \frac{(-1)^n}{(2n+1)!} \left\langle r^{2n} \right\rangle Q^{2n} .$$

Hence, r_p can be determined from

$$r_p \equiv \sqrt{\langle r^2 \rangle} = \left(-6 \left. \frac{\mathrm{d}G_E(Q^2)}{\mathrm{d}Q^2} \right|_{Q^2=0} \right)^{1/2} .$$

Of course, electron scattering cannot reach the exact $Q^2 = 0$ limit so in order to extract the charge radius from the data an extroplation must be made.

In the recent literature, you can find a number of extractions of the proton radius ranging from parsimonious fits of low Q^2 data [1–3] to extremely complex models with more than fifty free parameters [].

The criticism of the proponents of the complex fit is that the parsimonious models are biased, and implying that bias is some kind fundamental flaw that must be avoided. We will show in this work that when building a statistical model, one needs to in fact consider both bias and variance and, contrary to intuition, models with bias can in fact be the more appropriate model.

II. BIAS

A very straight forward example of bias being used to rule out simpler models can be found in a Z. Physik article from 1975 [4]; and while an older paper, its very strong conclusions are still noted to this day [5] and a similar exercises are done with other functions in Kruas*etal*. [6]. As noted in [7], the use of even an approximate generative model can be extremely important in understanding.

The example as problem presented in the Z. Physik paper is extremely simple to reproduce. Randomly generate sets of faux change form factor faux in steps of 0.05 fm^{-2} from 0.1 fm to 0.4, 0.8, 1.2 and 1.6 fm^{-2} using the standard dipole function:

$$\mathrm{G_D}(Q^2) = (1 + Q^2/(18.27 fm^{-2}))^{-2}. \qquad (1)$$

Perform fits on the resulting sets of faux data with linear

TABLE I. The following table shows the mean a_0 and radius terms from doing 1E6 Monte Carlo simulations for each range where Eq. 1 was used to generate faux data in 0.05 fm^{-2} steps with each points randomized using 0.5% normal distribution. The results clearly indicate that the linear fits are biased. The input radius was 0.8113 fm (an a1/a0 term of 0.1097 fm^{-1}) and an a0 of one.

interval	linear fit		quadratic fit	
fm^{-2}	a_0	radius	a_0	radius
0.1 – 0.4	1.000	0.79	1.000	0.81
0.1 – 0.8	0.999	0.78	1.000	0.81
0.1 – 1.2	0.997	0.77	1.000	0.81
0.1 – 1.6	0.996	0.76	1.000	0.81

and quadratic functions. The entire procedure is then repeated many times to determine the mean of the extracted radii for each model. Table II reproduces the table found in the Z. Physik article using python on a modern computer yielding only minor differences. As the table clearly shows, the mean of 1E6 linear fits is biased and thus the authors conclude that the linear models should be rejected in favor of the lower-bias quadratic function. They then proceed to extract the proton charge radius from real data using a five parameter fit: a quadratic charge form factor and three floating normalizations.

A Python notebook is included in the supplemental material.

III. VARIANCE

While the mean of the results is indeed correct; when we run an experiment we typically do no get to run it 1E6 times. In particular in nuclear physics, the experiments are few and far between thus we need to carefully consider variance as well as the bias when picking the statistical model to use.

Table III shows more complete picture of the simulation results where the variance is shown along with the bias. This table in fact shows nearly a textbook illustration of the trade-off between variance and bias with the simple fits having a relatively high bias with a low variance while the quadratic fits have a low bias and high variance.

IV. GOLDILOCKS DILEMMA

For any given statistical model, the goal is to find the optimal balance between bias and variance. In general, this can be written as:

$$\frac{dBias^2}{dComplexity} = \frac{-dVariance}{dComplexity} \tag{2}$$

Thus going back to Table III and checking the root mean square error, one can see that for the four ranges one finds the 0.1 – 0.8 range is actually optimal for the

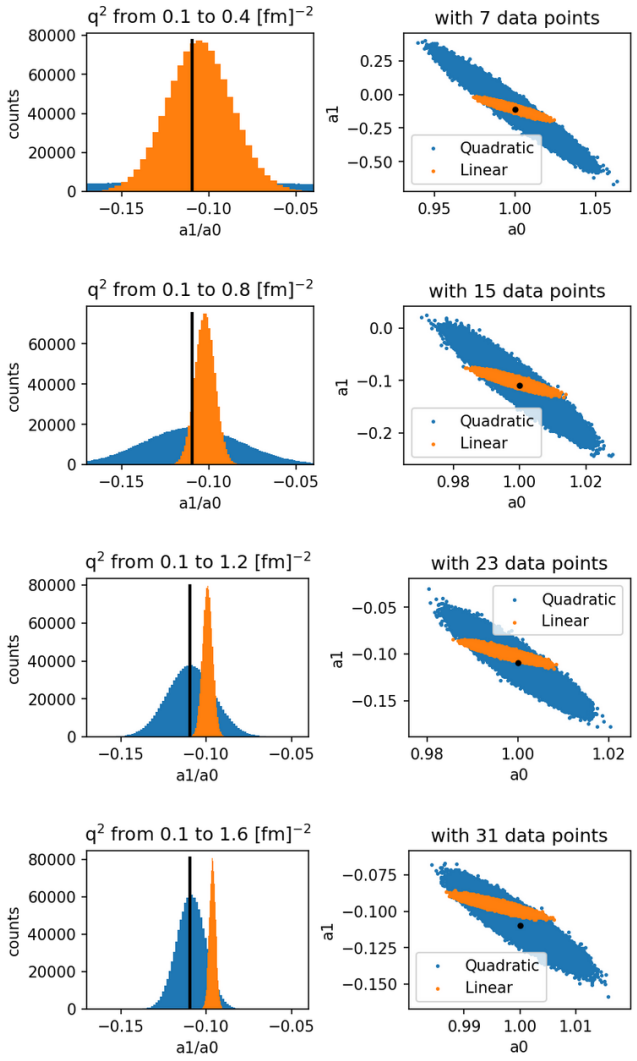


FIG. 1. Bla bla bla.

linear model and the 0.1 – 1.6 range is optimal for the quadratic model. This is in complete contrast to the conclusion one draws when only considers bias as presented in Table II though constant with the observation that the optimal specific form of the parameterization may depend on the Q^2 region being fit [**?**].

It is interesting to repeat the Monte Carlo simulation for equal number of data points within each range especially since, for any given experiment, the elastic scattering cross sections are significantly higher as lower values of Q^2.

As shown in Table III, the picture is even greyer as the root mean square error of the linear fit is nearly equal to the quadritic thus, assuming standard dipole was the true generating function, experiments with 31 data point and an uncertainty of 0.005 per point over a range of 0.1 to 0.8 and a different experiment over a range of 0.1 to 1.6 would produce nearly idenitical if all other things were equal.

TABLE II. The input radius was 0.8113 fm (an a1/a0 of 0.1097 fm^{-1}).

Data	Range	linear fit						quadratic fit					
Points	fm^{-2}	a0	Radius	a1/a0	Bias	Sigma	RMSE	a0	Radius	a1/a0	Bias	Sigma	RMSE
7	0.1 – 0.4	0.9995	0.7948	-0.1053	-0.0044	0.0184	0.0189	1.0000	0.8063	-0.1084	-0.0013	0.1094	0.1094
15	0.1 – 0.8	0.9987	0.7828	-0.1021	-0.0076	0.0057	0.0095	1.0000	0.8096	-0.1092	-0.0005	0.0281	0.0281
22	0.1 – 1.2	0.9975	0.7712	-0.0991	-0.0106	0.0030	0.0110	0.9999	0.8089	-0.1090	-0.0007	0.0138	0.0138
31	0.1 – 1.6	0.9959	0.7600	-0.0963	-0.0134	0.0019	0.0136	0.9998	0.8075	-0.1087	-0.0010	0.0085	0.0085

TABLE III. Same as before, but now with equal number of data points of each range.

Data	Range	linear fit						quadratic fit					
Points	fm^{-2}	a0	Radius	a1/a0	Bias	Sigma	RMSE	a0	Radius	a1/a0	Bias	Sigma	RMSE
31	0.1 - 0.4	0.9995	0.7951	-0.1054	-0.0043	0.0098	0.0107	1.0000	0.8090	-0.1091	-0.0006	0.0629	0.0629
31	0.1 - 0.8	0.9987	0.7829	-0.1021	-0.0076	0.0041	0.0086	1.0000	0.8099	-0.1093	-0.0004	0.0208	0.0208
31	0.1 - 1.2	0.9974	0.7712	-0.0991	-0.0106	0.0026	0.0109	0.9999	0.8089	-0.1091	-0.0006	0.0121	0.0121
31	0.1 - 1.6	0.9959	0.7600	-0.0963	-0.0134	0.0019	0.0136	0.9998	0.8076	-0.1087	-0.0010	0.0085	0.0085

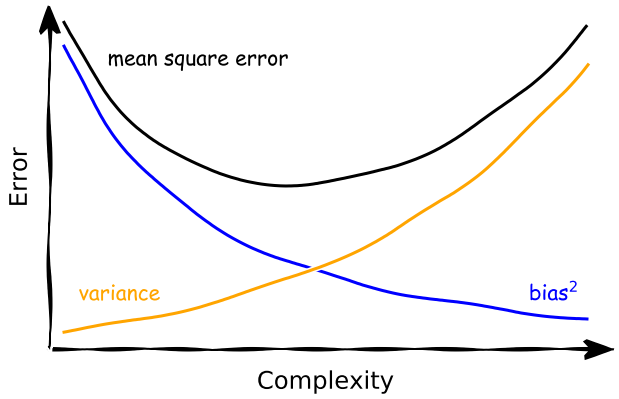


FIG. 2. An illustration of the trade-off between bias and variance when selecting a statistical model. Simple models will have low variance but high bias (under-fitting) while complex models will have low bias but high variance (over-fitting). It is this trade-off that one seeks to balance. While with repeated Monte Carlo simulations it is trivial to find the optimal predictive model for a give set of data; in the real world true model is typically unknown and one only gets preform a very limited number of experiments and thus one relys on using real data and statistical methods for model selection [8].

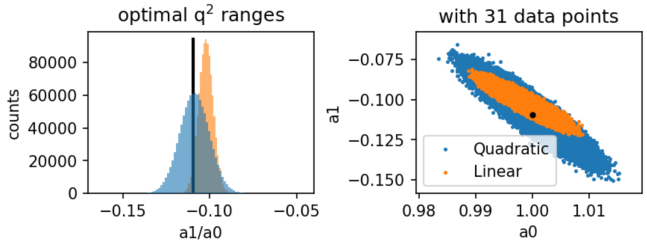


FIG. 3. Shown is the result of million simulations and fits of linear fits 0.1 – 0.8 fm^{-2} and quadratic fits 0.1 – 1.6 fm^{-2} both with 31 eqaully spaced data points. Using root mean square error as the matrix, neither example is significantly better then the other for exacting the proton radius This is annaligous to a dart game between two eqaully skilled players though one who hits the bulls eye more often yet has a large spread (low bias but high variance) and another equally skilled player who has a tigher cluster of hits but an offset (high bias but low variance).

In this case, the choice of the parsimonious modeler to use the low Q^2 data would like be driven by the recugnition that as Q^2 increases the extraction of a charge form factor is complicated by the growing influcance of the magnetic form factor while the use of the larger Q^2 range would likely be driven by a desire to form a more complet picture of the proton's structure (e.g. interest not only in the proton's radius but also its higher order momentums).

$$\sigma/\sigma_{Mott} = epsG_E^2 + tauG_M^2 \tag{3}$$

The above illistration in fact suggests there are indeed two path forward for the modeler. Either use simple models and extremely low Q2 data or use more complex models that cover larger ranges of data. It is worth noting that inorder to use larger ranges of experimental data, one will need to model not only the charge form factor but also the exact details of magnetic form factor will start become quite important.

V. MODEL SELECTION AND THE MODERN DEBATE

While this classic Monte Carlo example problem is over 40 years old, it actually points to exactly the split in the current electron scattering proton radius extraction proceedures. The parsimonious modelers who are focused solely on extracting a radius have focused in on the low Q^2 region accepting accepting a higher level of bias in exchange for low variance while those modelers who are interesting in extracting more information about the proton (i.e. higher order moments) fit longer Q^2 ranges and

have focused on complex models which while lower in bias come at a cost of higher varaince. And the end result has been an uncertainty on the electron scattering proton radius stuck at $\approx$. 0.01 fm since L. Hand *et al.* original fit in 1963 [] as the increased complexity drives an increased variance.

Also, since we don't know the true model so one cannot exactly calculate the RMSE; so while exercises like this one are extremely useful for making sets are models; in the end, the data must be used to select the approximate model. For this we relay on statistical modeling techinques such as chi2, reduced chi2, F-tests, A.I.C., B.I.C. to guide our selecton of regression models and/or theoretical constraints.

In fact, statistics books warn about drawing too strong an inference from these types of Monte Carlos. For example, just because the linear model has a negative bias when compared to standard dipole, does not imply that it has a negative bias to all possible models. In fact, the Z.Physik paper itself rules out the very model it was using for model builting: i.e. drawing very strong conclusions from the standard dipole function with its 0.81 fm radius yet their five parameter fit (two charge form factor parameters and three normalization parameters) gave a radius of 0.87 fm.

Looking back to our example function, one can see that the range of the Saskatoon data (0.1 – 0.8 fm^{-2}) could in fact be a reasonable choice for extracting a radius and as shown much later if one redoes the Z.Physik fits but adopts the Saskatoon linear proceedure one finds a radius of 0.84 fm [3]. Though want one really wants is a function with both low bias and low variance.

VI. BEYOND STANDARD DIPOLE AND POLYNOMIAL FUCNTIONS

While one can cointue to increase the Q^2 range of the data, one will find that higher and higher order polynomials are required to discribe the data and as shown in [6], but this leads to ever increasing variance and instability as real data isn't prefectly normally distriubuted as in our example Monte Carlo.

To try to avoid these issues, one can impliment a bounded least squares fit [] or try a machine learning regression techinque such as stepwise regresson to pick the appropriate order fit, though it would be far more statifying to simply find function that has both low bias and low varaince over a large range of Q^2.

While a Taylor series is the most general minimax function, it is not the only one. If one explores (quote math book) one will find several other functions which are equally flexible yet have properties that may make them approxiate for the problem at hand.

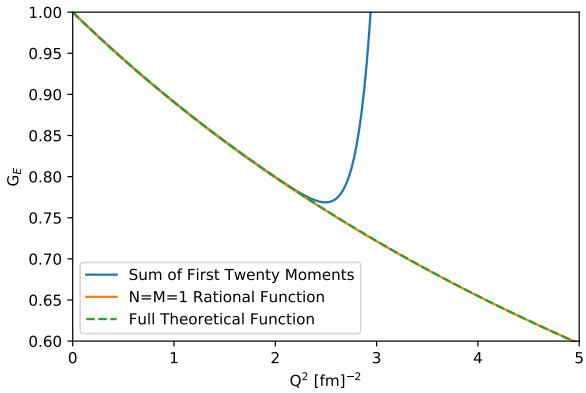
The rational function

$$G_E = (a_0 + a_1 Q^2)/(1 + a_2 Q^2) \tag{4}$$

And then a table of the results.

Add function of Jose and Christian.

$$\begin{aligned} TrailMoments = r0 &:1, \\ r2 &:0.71, \\ r4 &:1.4886245922111714, \\ r6 &: 8.396382952136717, \\ r8 &: 125.4944297151659, \\ r10 &: 3909.7337848855677, \\ r12 &: 203387.02342771934, \\ r14 &: 1.5533472343250945e7, \\ r16 &: 1.620661306754837e9, \\ r18 &: 2.2059826827872018e11, \\ r20 &: 3.790434839036413e13, \\ r22 &: 8.016789155142601e15, \\ r24 &: 2.0455477973159212e18, \\ r26 &: 6.194020562555066e20, \\ r28 &: 2.1954425776105903e23, \\ r30 &: 9.003115066949057e25, \\ r32 &: 4.228910288487866e28, \\ r34 &: 2.255488525203844e31, \\ r36 &: 1.3555113752177442e34, \\ r38 &: 9.117417045801919e36, \\ r40 &: 6.822151726705506e39 \end{aligned}$$



$TrailG_E(Q^2) =$

$1.0 * q2 * *0+$
$-0.11833333333333333 * q2 * *1+$
$0.012405204935093095 * q2 * *2+$
$-0.0016659489984398247 * q2 * *3+$
$0.0003458290060492887 * q2 * *4+$
$-9.794707453717652e-05 * q2 * *5+$
$3.26620112506641e-05 * q2 * *6+$
$-1.187870063325348e-05 * q2 * *7+$
$4.556419987712978e-06 * q2 * *8+$
$-1.8134578995563369e-06 * q2 * *9+$
$7.418995770908466e-07 * q2 * *10+$
$-3.1010304674158173e-07 * q2 * *11+$
$1.3187544953627 91e-07 * q2 * *12+$
$-5.6883963856478145e-08 * q2 * *13+$
$2.4830374072102887e-08 * q2 * *14+$
$-1.0948913684904526e-08 * q2 * *15+$
$4.87015501923618e-09 * q2 * *16+$
$-2.182769986978775e-09 * q2 * *17+$
$9.848413919033538e-10 * q2 * *18+$
$-4.469786132456496e-10 * q2 * *19+$
$2.039353202104902e-10 * q2 * *20+$
...

We note that this deceptively simple form can be extended to give the correct asymmotic behiver as was done by Kelly []

VII. APPLICTION TO EXPERIMENTAL DATA

We use the entire Mainz spectrometer B data set.

Results:

Pade Output Proton Charge Radius 0.8505217414548519
Pade Output Proton Magnetic Radius 0.8664383622994243

Total Chi2 = 679.35 for 681 data points, Chi2/v = 1.00

Checking just the lowest beam energy:

180 MeV: Total Chi2 176.629 wth 166 points fitting for a chi2/v 1.077:

VIII. SUMMARY

The concept of a bias-varinace trade-off is key to regulariztion techniques such as stepwise regression, ridge regression and statistical lasso. By accepting some bias, these techniques tend to achieve a far superior mean square error then the unregulated ordinary least squares solution of the same complexity.

For the specific example of electron scattering, we have shown that the practice of simply concluding that a model with a higher predictive validity is truer is not valid assumption and that parsimonious modeling can in fact have the higher predictive validity depending on the exact range and precision of the data.

The detailed math dehind very specific example given herein is covered in detail in Shmueli's review article [9]. And as George Box was always quick to say, all models are wrong; but some are useful.

[1] K. Griffioen, C. Carlson, and S. Maddox, Phys. Rev. **C93**, 065207 (2016), arXiv:1509.06676 [nucl-ex].
[2] M. Horbatsch, E. A. Hessels, and A. Pineda, Phys. Rev. **C95**, 035203 (2017), arXiv:1610.09760 [nucl-th].
[3] D. W. Higinbotham, A. A. Kabir, V. Lin, D. Meekins, B. Norum, and B. Sawatzky, Phys. Rev. **C93**, 055207 (2016), arXiv:1510.01293 [nucl-ex].
[4] F. Borkowski, G. G. Simon, V. H. Walther, and R. D. Wendling, Zeitschrift für Physik A Atoms and Nuclei **275**, 29 (1975).
[5] I. Sick and D. Trautmann, Phys. Rev. **C95**, 012501 (2017), arXiv:1701.01809 [nucl-ex].
[6] E. Kraus, K. E. Mesick, A. White, R. Gilman, and S. Strauch, Phys. Rev. **C90**, 045206 (2014), arXiv:1405.4735 [nucl-ex].
[7] D. W. Hogg, J. Bovy, and D. Lang, (2010),

arXiv:1008.4686 [astro-ph.IM].
[8] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning* (Springer, New York, 2009).
[9] G. Shmueli, Statist. Sci. **25**, 289 (2010).