Applying Statistical Modeling Techniques To The Proton Radius Puzzle

Douglas W. Higinbotham (Jefferson Lab)

along with
Vincent Lin (Western Branch High School / Rodman Scholar at University of Virginia)
David Meekins and Brad Sawatzky (Jefferson Lab)
Al Amin Kabir (Kent State)
Blaine Norum (University of Virginia)
Celina Pearson (Virginia Tech)
Ethan Buck, Carl Carlson & Keith Griffioen (William & Mary)
With statistics advice from Simon Širca (Ljubljana) and Sanjoy Mahajan (MIT/Olin College)

What's to know?

Name	Statistic
chi-squared distribution	$\sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$

Just fit until you get $\chi^2 / \nu = 1$ and your good? Right…. ?!

(where ν is the degrees of freedom in the fit)

What could possibly go wrong?!

What if the weights (sigma's) are underestimated or overestimated?
What if I have the wrong model?
What if the data aren't normally distributed?
What if average redcued χ^2 is good, but one over-fits one area and under-fits another!!
(It is not trivial and just getting a reduced χ^2 ~ 1 does not mean you have a good result.)

All Models Are Wrong

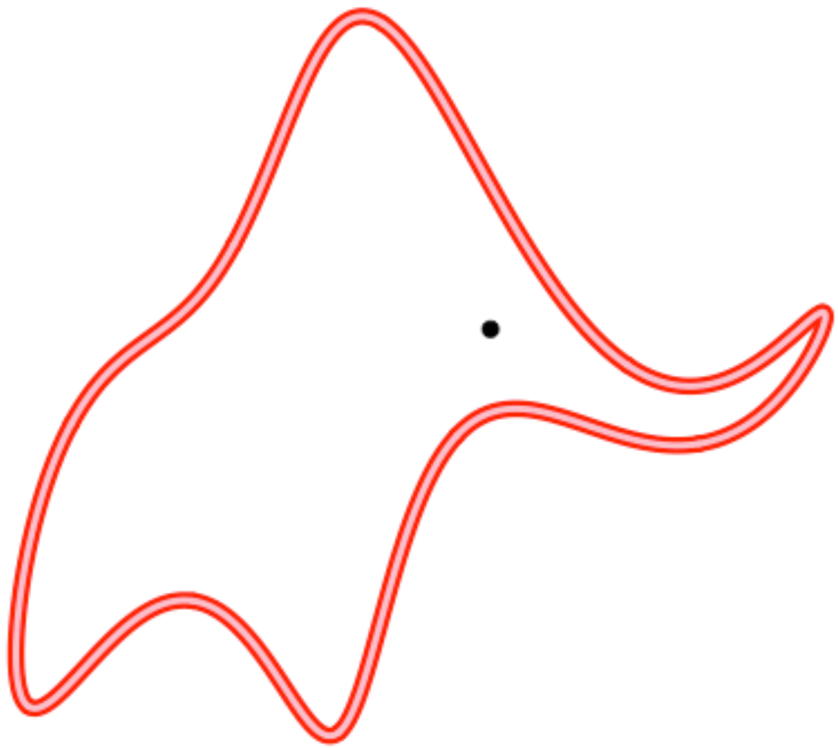
"The most that can be expected from any model is that it can supply a useful approximation to reality: All models are wrong; some models are useful." - George Box (1919 – 2013)

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk. " John von Neumann (1903 – 1957)

Freeman Dyson presents his model to Enrico Fermi: http://webofstories.com/play/4402

The Five Parameter Elephant

"Drawing an elephant with four complex parameters"
by Jurgen Mayer, Khaled Khairy, and Jonathon Howard, Am. J. Phys. 78 (2010) 648.



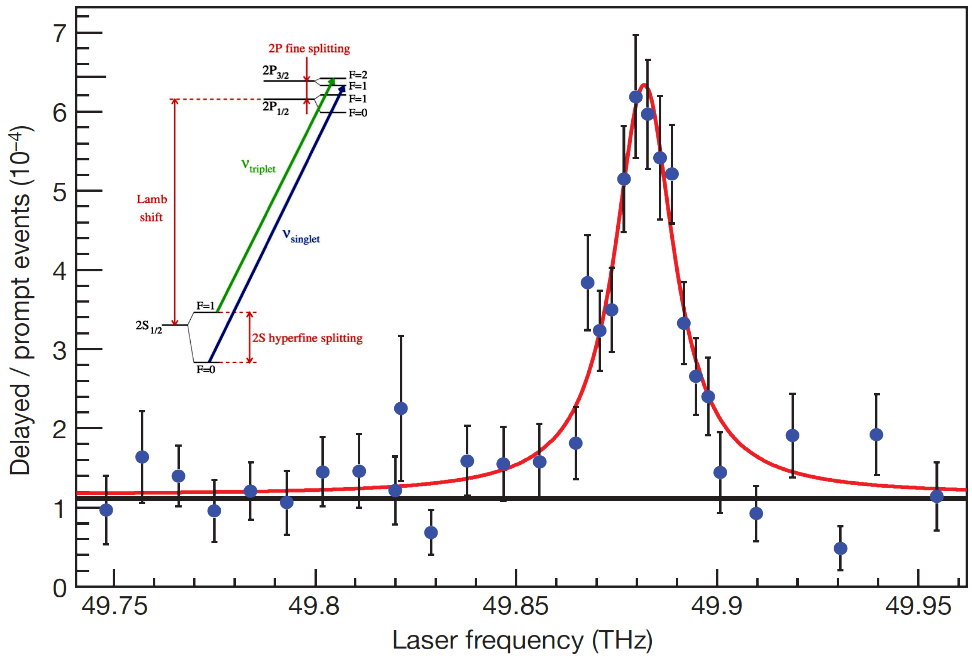
https://www.johndcook.com/blog/2011/06/21/how-to-fit-an-elephant/

Occam's Razor

- William Occam (1287 – 1347)
- One can always explain failing explanations with an ad hoc hypothesis, thus in Science, simpler theories are preferable to more complex ones. (e.g. the Sun centered vs. Earth centered)
- Layman's version of Occam's Razor is "the simplest explanation is usually the correct one" (i.e. KISS)
- In statistical versions of Occam's Razor, one uses a rigorous formulation instead of a philosophical argument. In particular, one must provide a specific definition of simple:
  - F test, Akaike information criterion, Bayesian information criterion, etc.
  - **In statistical modeling of data too simple is under-fitting and too complicated is over-fitting.**

Muonic Hydrogen Data

- High precision results from Muonic Lamb shift data give a proton radius of 0.84 fm.
- This result contradicts many other extractions which have determined the proton radius to be ~0.88 fm.



$$E_{2p} - E_{2s} = 209.98 - 5.2262\, r_p^2 + 0.0347\, r_p^3 \text{ meV}$$

NOTE: The radius in this formula is consistent with other extractions.

Some Of The Possible Explanations

Lamb shift difference of 310 ± 2 μeV

μH experiment wrong ?

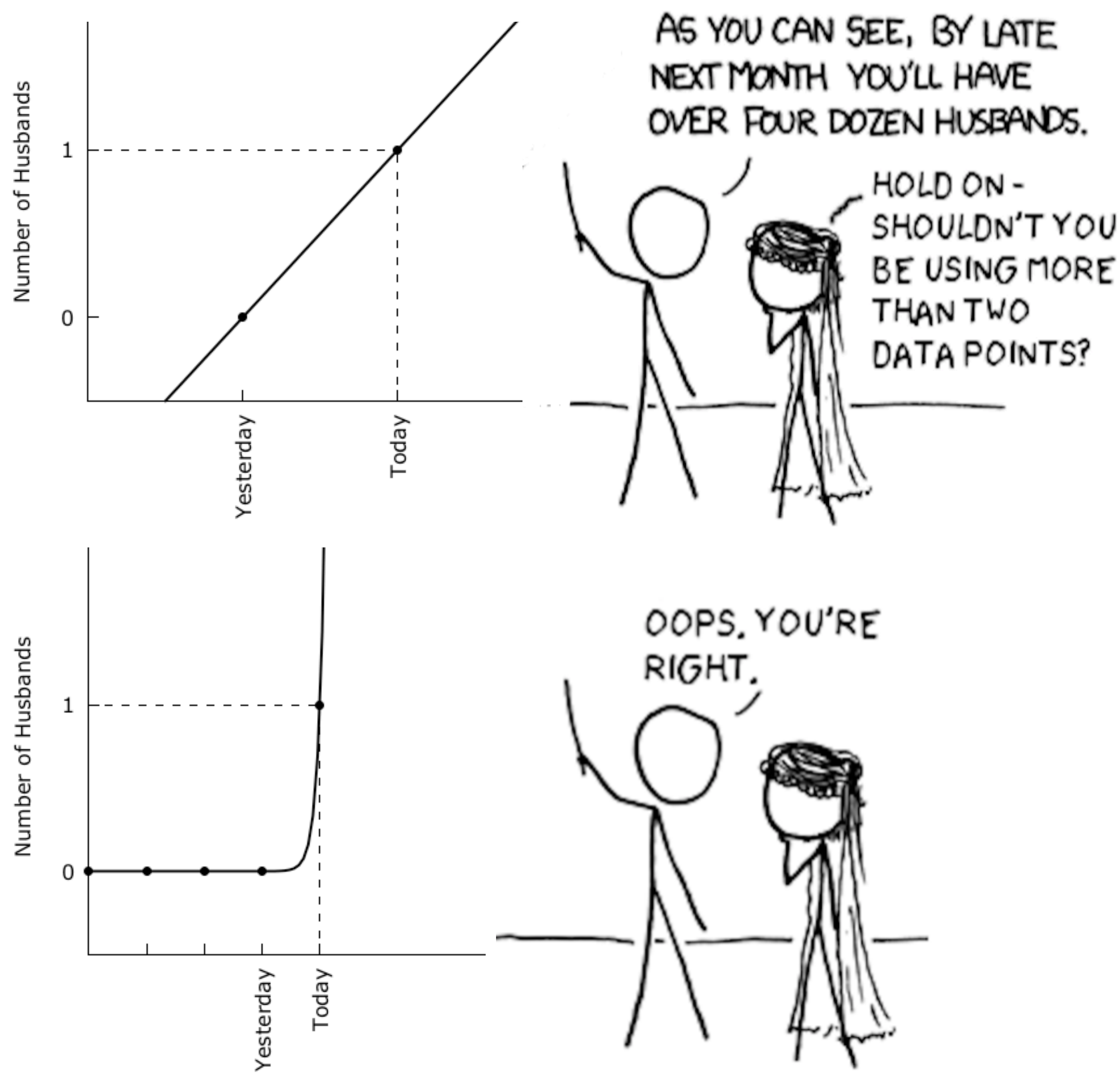
eH theory wrong ?

Beyond Standard Model

μH theory wrong ?
- 2γ corrections
- missing QED or EW corrections
- soft hadronic corrections

eH experiment + ep scattering wrong ?
- R_∞
- 2γ corrections
- low-Q^2 extrapolation

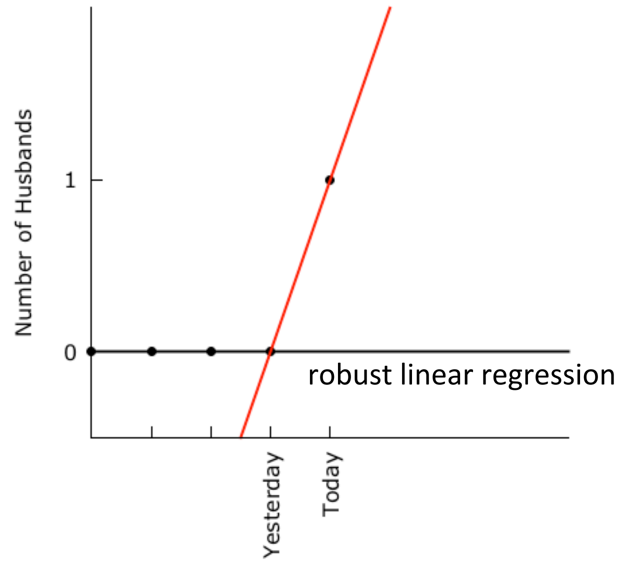
XFCD “My Hobby: Extrapolating”



GNUPLOT OVERFITTING CODE
Using 101,600 Iterations To Converge

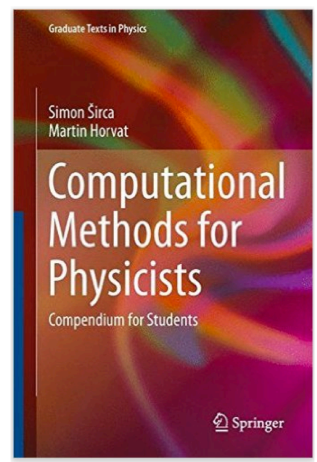
```
#
# gnuplot overfitting of xkcd Husband Data
# modified from https://xkcd.com/605/
# by
# Douglas W. Higinbotham
#

set terminal wxt  enhanced font "verdana,12" size 900,450
set nokey
set xtic rotate 90
set ytic 0,1,1
set border 3

set xtics nomirror
set ytics nomirror

set multiplot layout 1,2

set ylabel "Number of Husbands"

f(x)=f0+f1*x
g(x)=g0*exp(g1*x)

fit f(x) '1.dat' using 1:3 via f0,f1
fit g(x) '2.dat' using 1:3 via g0,g1


set arrow from 0,1 to 2,1 nohead dashtype 7 lc 'black'
set arrow from 2,-0.5 to 2,1 nohead dashtype 7 lc 'black'
set xrange [0:3]
set yrange [-0.5:2]
plot '1.dat' using 1:3:xtic(2) lt 7 lc 'black' ,f(x) lw 2 lc 'black'
unset arrow

set xrange [-2:6]
set arrow from -2,1 to 2,1 nohead dashtype 7 lc 'black'
set arrow from 2,-0.5 to 2,1 nohead dashtype 7 lc 'black'
plot '2.dat' using 1:3:xtic(2) lt 7 lc 'black' ,g(x) lw 2 lc 'black'

unset multiplot
pause -1
```

Robust Linear Regression



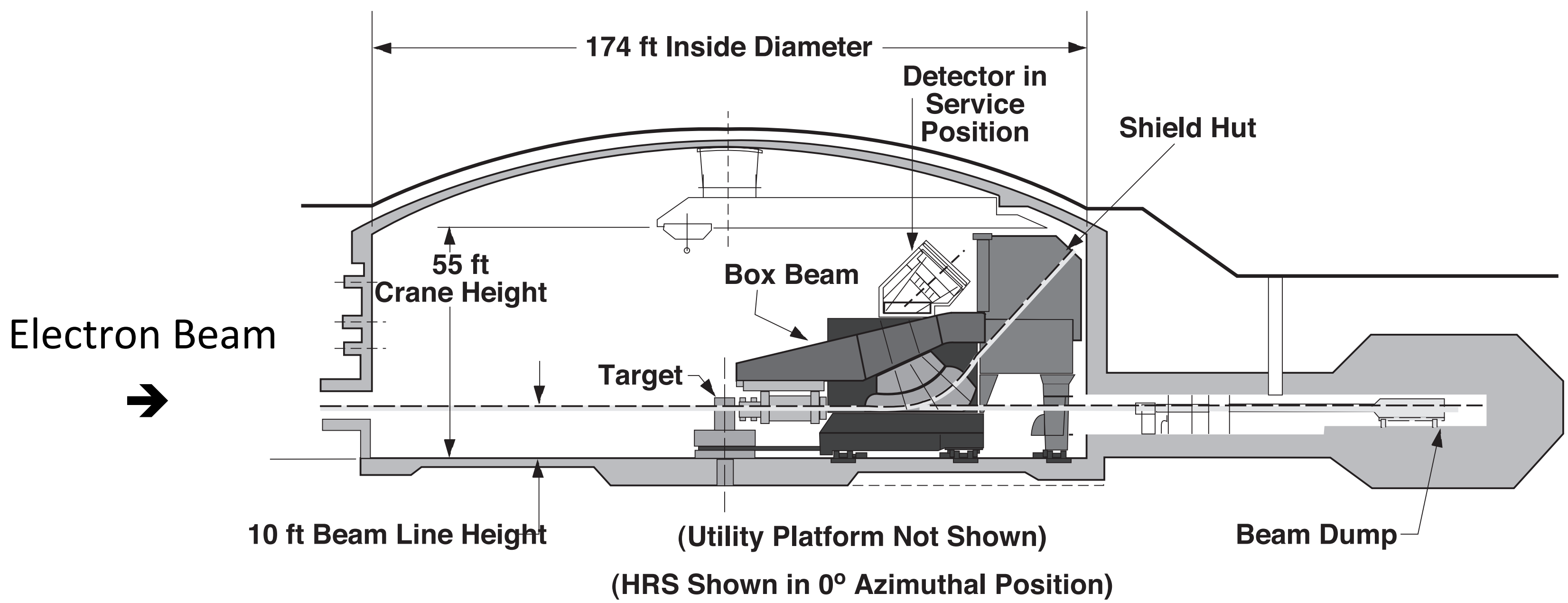
"An ever increasing amount of computational work is being relegated to computers, and often we almost blindly assume that the obtained results are correct."

- Simon Širca & Martin Horvat



How do we make the electron scattering measurements?

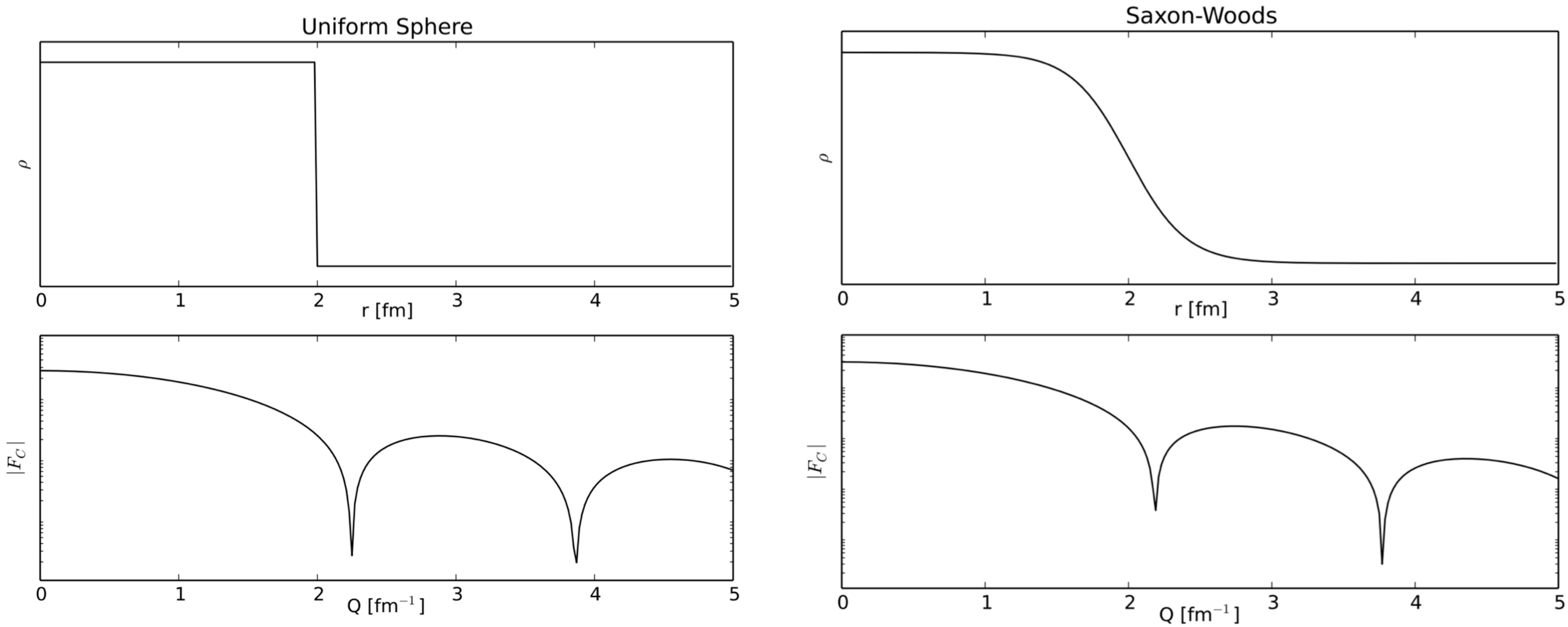
- Beam of electrons from an accelerator (E)
- Place target material in the beam
  - Foils are easy, nearly point (typically thin) targets and thickness is easy to determine
  - Cryo-targets are challenging (e.g. boiling effects, energy loss)
- For elastic hydrogen measure scattered electron (E') and/or proton.
  - Over determined reaction
- Spectrometers are used
  - Magnetic fields, wire-chambers, reconstructed tracks, sieve data, etc.



Jefferson Lab Hall A Left Spectrometer

Electron Scattering Charge Radii from Nuclei

Fourier Transformation of Ideal Charge Distributions.



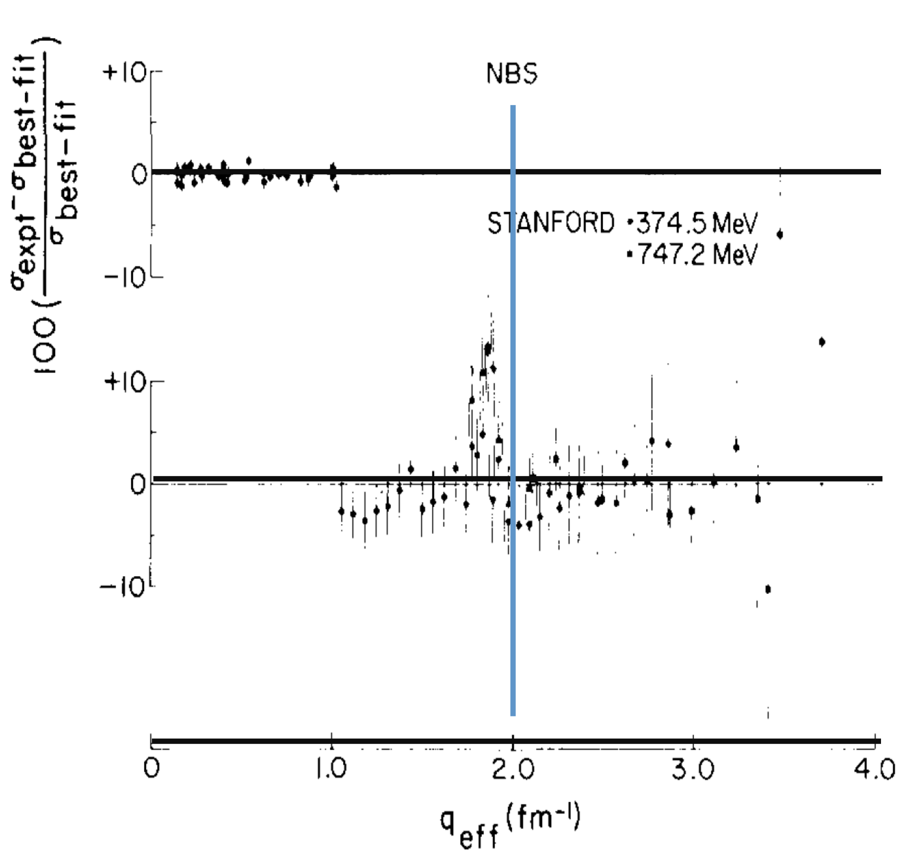
Example Plots Made By R. Evan McClellan (Jefferson Lab Postdoc)

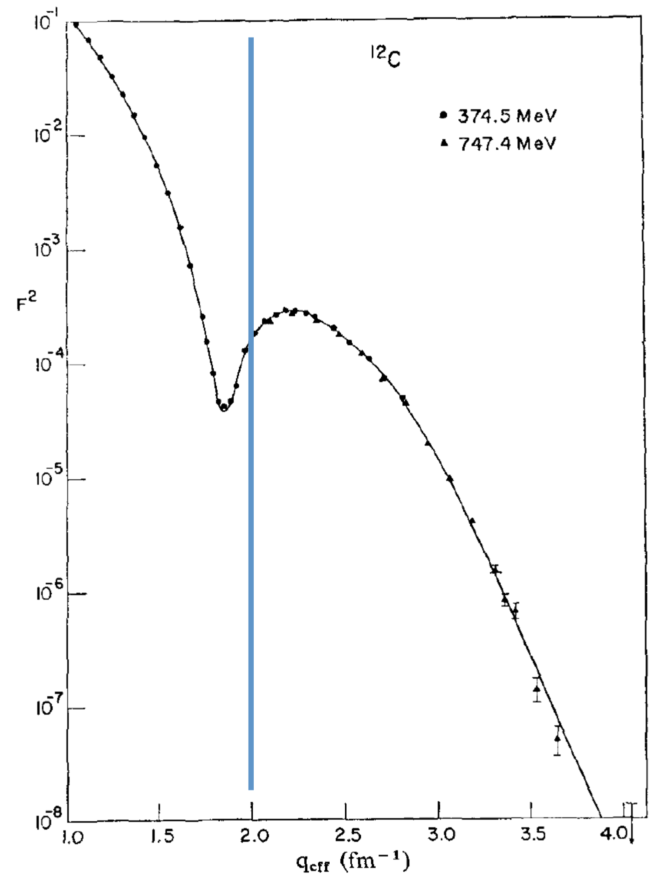
e.g. for Carbon: Stanford high Q^2 data from I. Sick and J.S. McCarthy, Nucl. Phys. **A150** (1970) 631.
National Bureau of Standards low Q^2 data from L. Cardman *et. al*., Phys. Lett. **B91** (1980) 203.

Determining the Charge Radius of Carbon

Stanford high Q^2 data from I. Sick and J.S. McCarthy, Nucl. Phys. **A150** (1970) 631.

National Bureau of Standards (NBS) low Q^2 data from L. Cardman et. al., Phys. Lett. **B91** (1980) 203.





See the L. Cardman's paper for details of the carbon radius (2.46 fm) analysis.

Charge Radius of the Proton

- Proton G_E has no measured minima and it is too light for the Fourier transformation to work in a model independent way.
- Thus for the proton we make use of the fact that as Q^2 goes to zero the charge radius is proportional to the slope of G_E

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

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

We don't measure to Q^2 of zero, so this is going to be an extrapolation problem.
NOTE: There is general agreement that this definition of r_p is consistent with the Muonic results.

Elastic Electron Scattering

From relativistic quantum mechanics one can derive the formula for electron-proton scattering where one has assumed the exchange of a single virtual photon.

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_{\mathrm{Mott}} \cdot \frac{E'}{E}\left[\frac{G_E^2 + \tau G_M^2}{1+\tau} + 2\tau G_M^2 \tan^2\frac{\theta}{2}\right]$$

where G_E and G_M form factors take into account the finite size of the proton.

$G_E = G_E(Q^2)$, $G_M = G_M(Q^2)$; $G_E(0)=1$, $G_M(0) = \mu_p$

$Q^2 = 4\,E\,E'\sin^2(\theta/2)$ and $\tau = Q^2/4m_p^2$

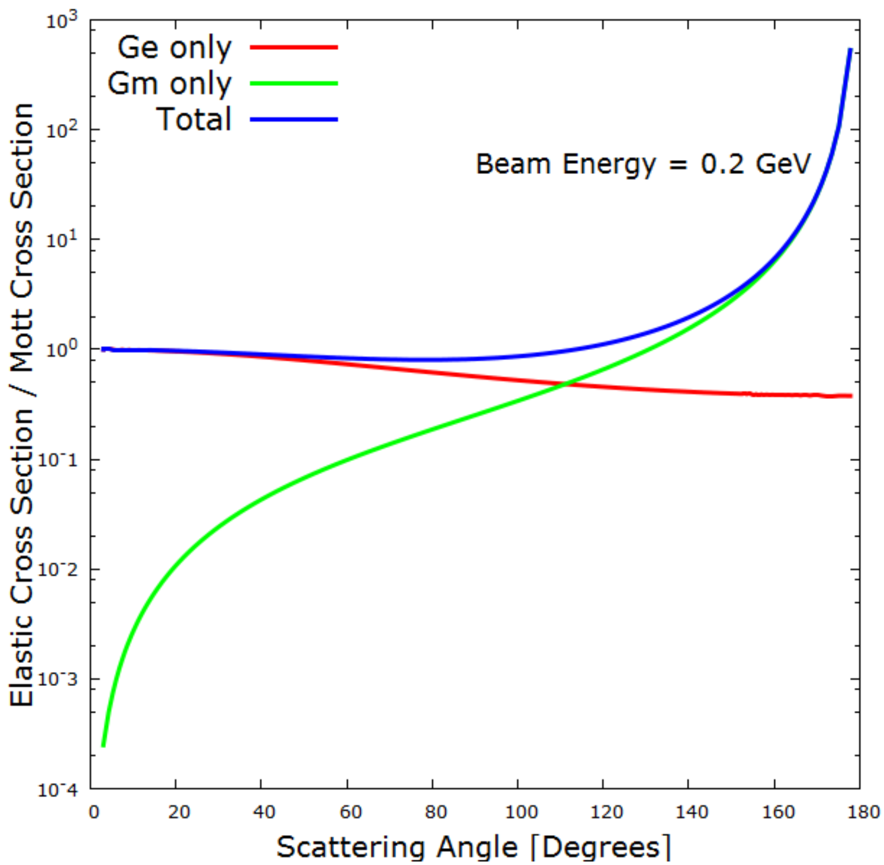
Elastic cross sections at small angles and small Q^2's are dominated by G_E (JLab PRad Hall B)

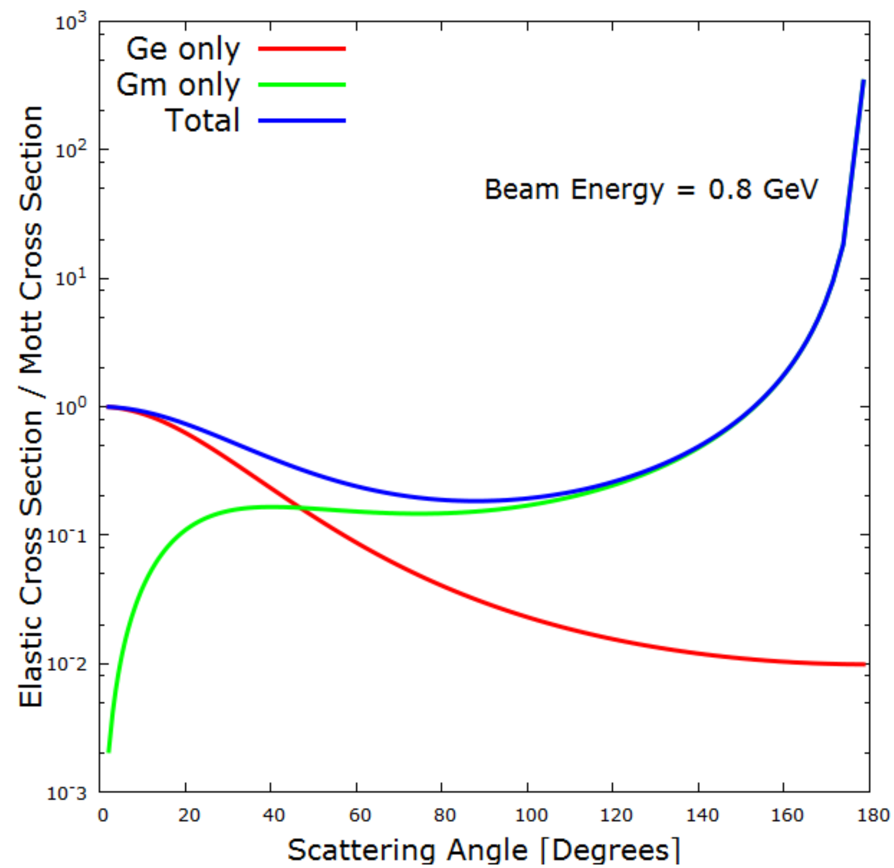
Elastic cross sections at large angles and large Q^2's are dominated by G_M (JLab GMP Hall A)

For moderate Q^2's one can separate G_E and G_M with the Rosenbluth technique (same Q^2 different E,θ).

G_E and G_M Contributions To The Cross Section

Plots by Ethan Buck (Jefferson Lab SULI Student and W&M undergraduate)

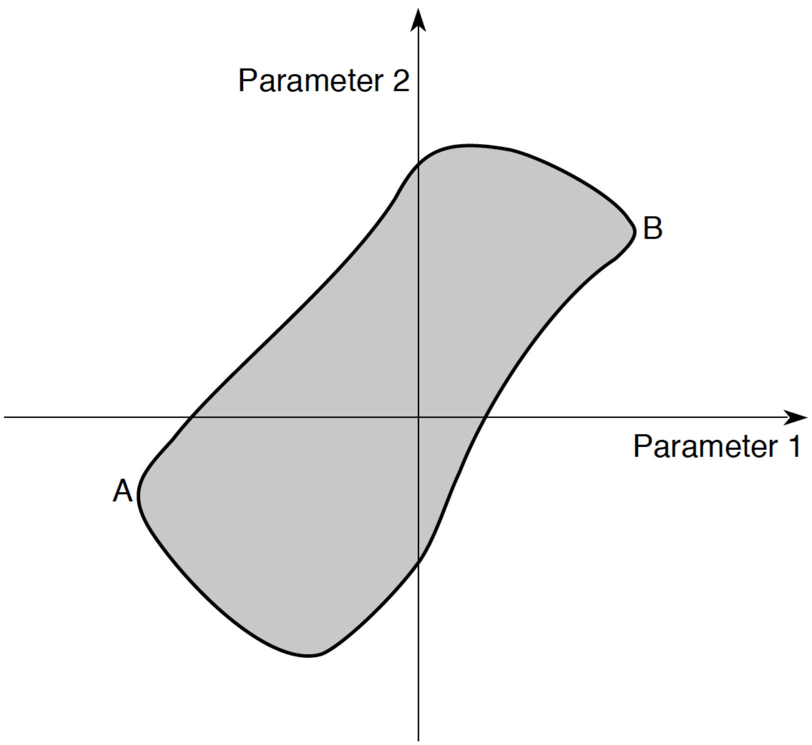




Experiments like PRad (Hall B) go to small angle to maximize G_E and minimize G_M contribution..

Global fits, like typically done with the Mainz 2010 data, need several normalization, G_E and G_M

Multivariate Errors



The Interpretation of Errors in Minuit (2004 by James)

seal.cern.ch/documents/minuit/mnerror.pdf

As per the particle data handbook, one should be using a co-variance matrix and calculating the probably content of the hyper-contour of the fit. Default setting of Minuit of "up"(normally called $\Delta\chi^2$) is one.

Also note standard Errors often underestimate true uncertainties. (manual of gnuplot fitting has an explicate warning about this)

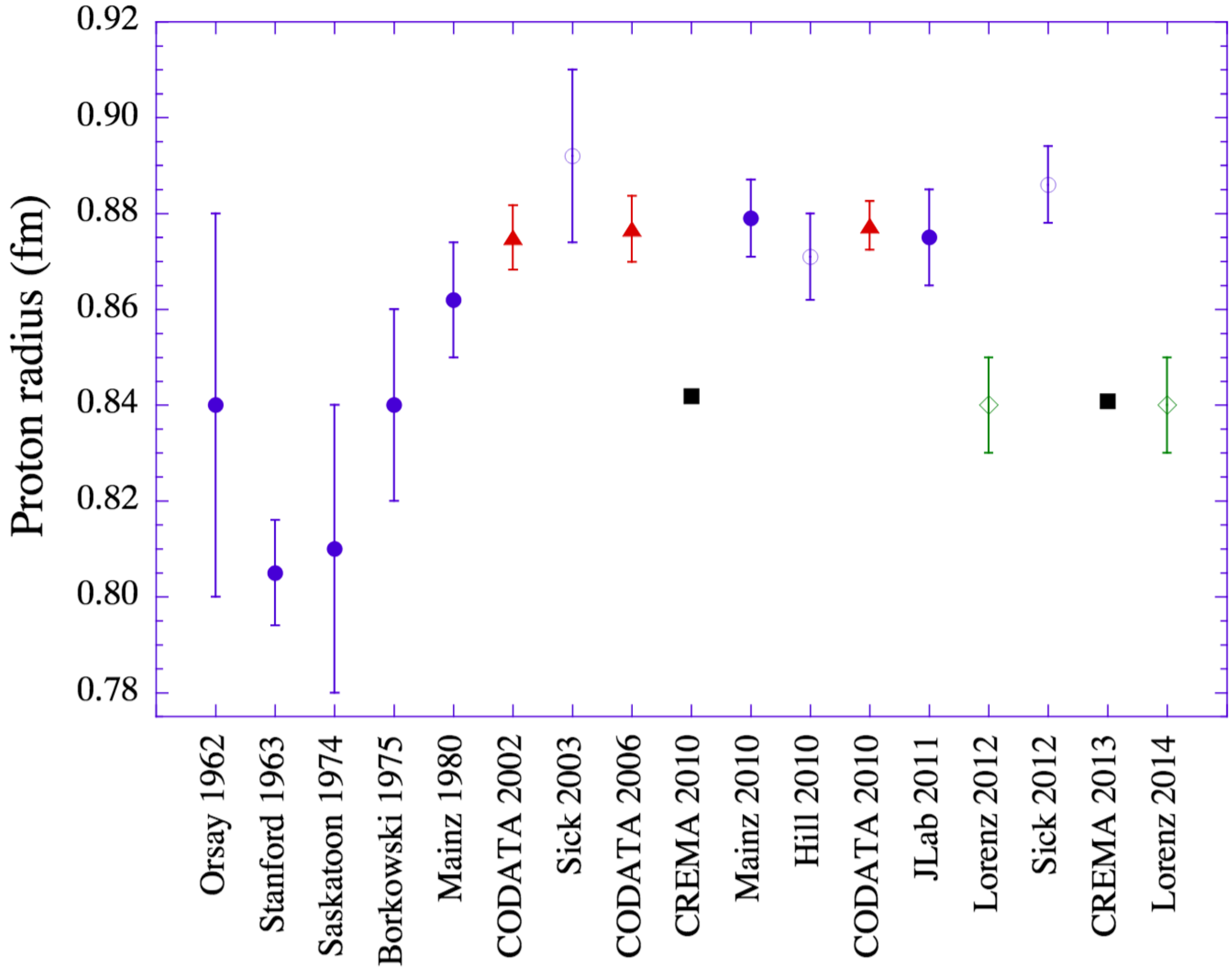
Number of Parameters	Confidence level (probability contents desired inside hypercontour of $\chi^2 = \chi^2_{\min} + \mathrm{up}$)				
	50%	70%	90%	95%	99%
1	0.46	1.07	2.70	3.84	6.63
2	1.39	2.41	4.61	5.99	9.21
3	2.37	3.67	6.25	7.82	11.36
4	3.36	4.88	7.78	9.49	13.28
5	4.35	6.06	9.24	11.07	15.09
6	5.35	7.23	10.65	12.59	16.81
7	6.35	8.38	12.02	14.07	18.49
8	7.34	9.52	13.36	15.51	20.09
9	8.34	10.66	14.68	16.92	21.67
10	9.34	11.78	15.99	18.31	23.21
11	10.34	12.88	17.29	19.68	24.71
	If FCN is $-\log(\text{likelihood})$ instead of χ^2, all values of up should be divided by 2.				

In ROOT: **SetDefaultErrorDef(real #)**

Default is 1 and doesn't change unless you change it!

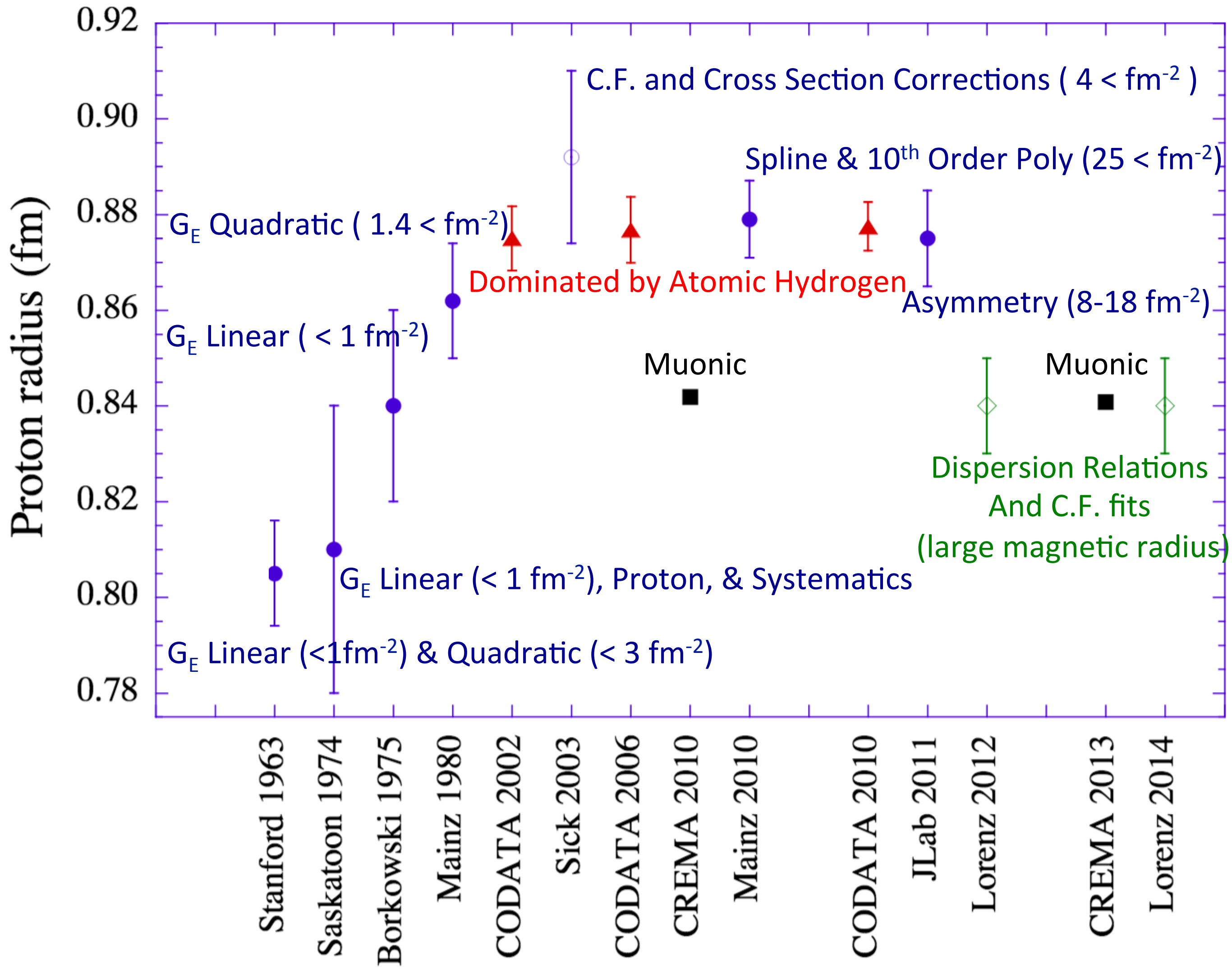
Proton Radius vs. Time

V. Punjabi et al., Eur. Phys. J. **A51** (2015) 79.



And Ever Changing Fit Functions

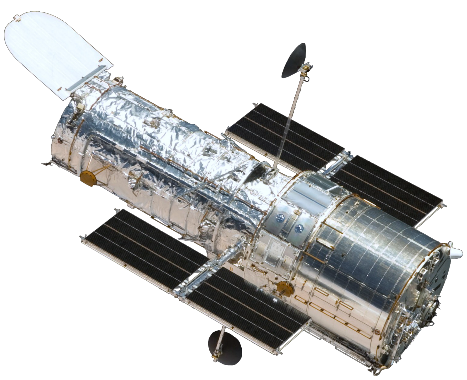
V. Punjabi et al., Eur. Phys. J. **A51** (2015) 79.



Measurement Is Often A Goldilocks Problem

From Deep Space

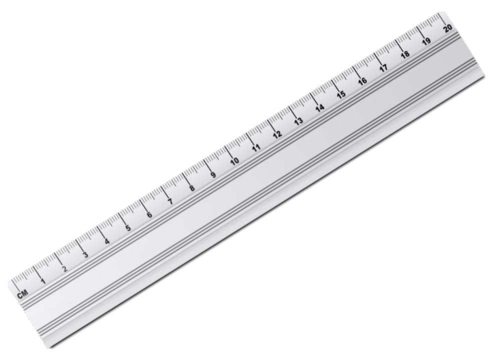
Too Far



A Modern Telescope

From Orbit

Just Right



Ruler & Some Geometry

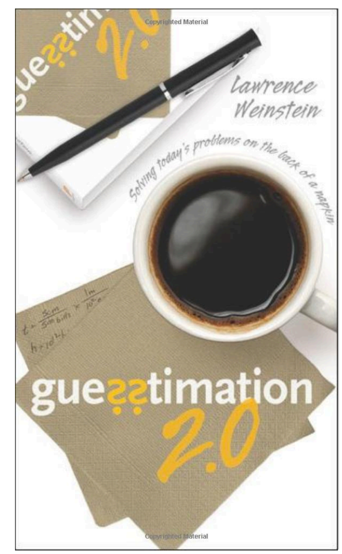
On The Planet

Too Close

Theodolite*

What is *just right* for the proton?!

- We use **Plank's constant** one to relate energy to length in natural units:
  - **Q^2 of 1 GeV^2 = 25.7 fm^{-2}.**
- Radius of the proton is ~ 0.84 - 0.88 fm
- Thus one can immediately guesstimate that with electron scattering one needs:
  - $Q^2 < (1/0.88\ fm)^2 < 1.2\ fm^{-2}$ to get the radius of the proton (equivalent to 0.05 GeV^2)
  - $Q^2 > 1.2\ fm^{-2}$ to understand the details of the edge of the proton (e.g. a pion cloud, CQCBM, etc.)
  - $Q^2 >> 1.2\ fm^{-2}$ to understand transition from hadronic to partonic (e.g. the bound light constitute quarks)



Guesstimation books by Larry Weinstein (ODU)

Extrapolate The Slope of G_E Using Low Q^2 Data

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

The question is going to be what function to use for the fitting & extrapolating.

The answer to this question STRONGLY effects the answer!

For linear regression using a polynomial function one can use an F-test.
For non-linear regression more complicated techniques are required.

Warning: Danger of Confirmation Bias

In psychology and cognitive science, confirmation bias is a tendency to search for or interpret information in a way that confirms one's preconceptions, leading to statistical errors.

Test of Additional Term

A textbook statistics problem is to quantify when to stop adding terms to a fit of experimental data.

One way to do this is with an F-distribution test.

$$F = \frac{\chi^2(j-1) - \chi^2(j)}{\chi^2(j)}(N-j-1)$$

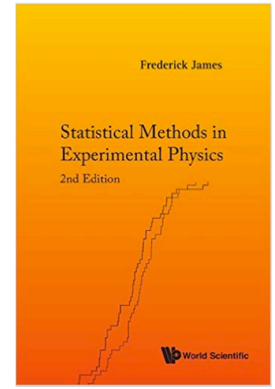
where j is the order of the fit and N the number points being fit.

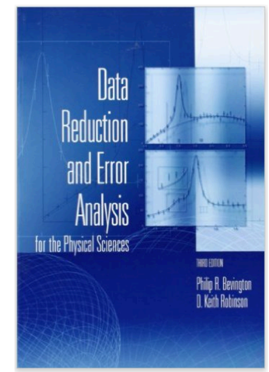
$N-j-1$	2	3	4	6	8	12	20	60	120
Reject j^{th} order to 95% confidence level if F is smaller than	18.5	10.1	7.7	6	5.3	4.7	4.3	4	3.9

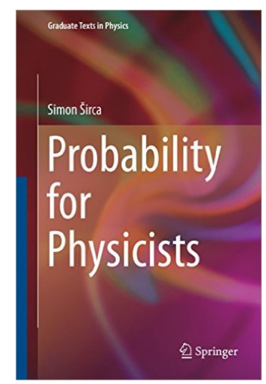
Quantifies a statement that adding a term doesn't significantly improve a fit.

One is free to pick a different alpha, alpha=0.05 is just typical to prevent over-fitting.

(see James 2nd edition page 282, Bevington 3rd edition page 207, or Širca page 268)

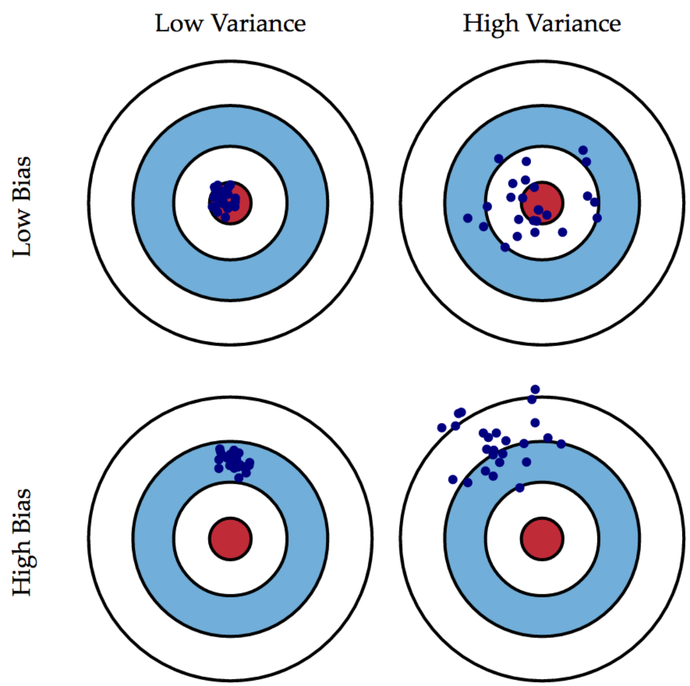


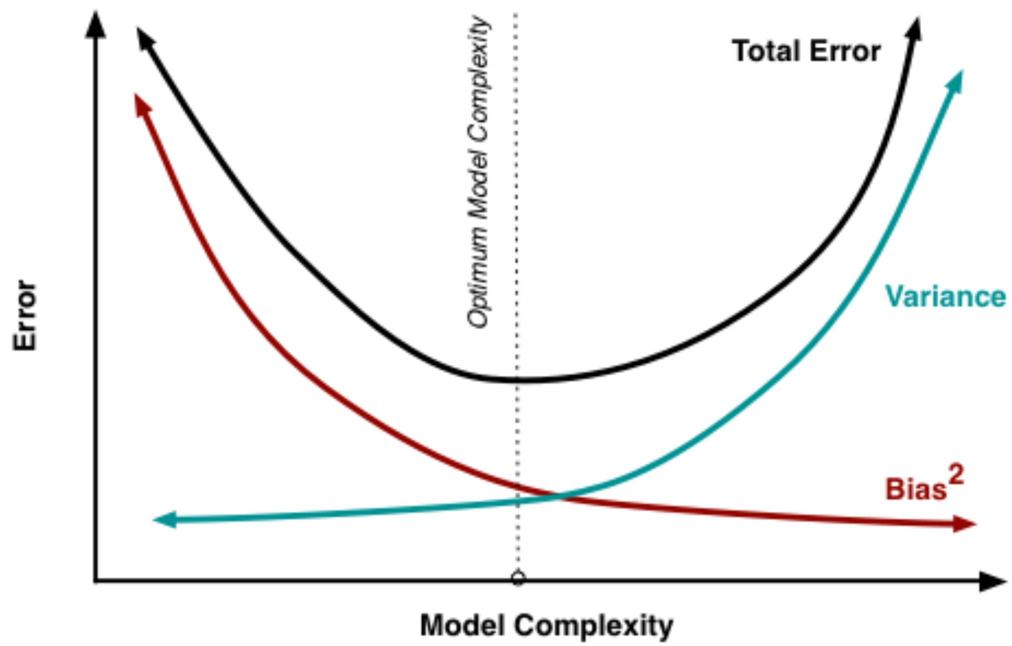




Bias vs. Variance

http://scott.fortmann-roe.com/docs/BiasVariance.html





"Models with low bias are usually more complex (e.g. higher-order regression polynomials), enabling them to represent the training set more accurately. In the process, however, they may also represent a large noise component in the training set, making their predictions less accurate - despite their added complexity. In contrast, models with higher bias tend to be relatively simple (low-order or even linear regression polynomials), but may produce lower variance predictions when applied beyond the training set."

NOTE: We run ONE experiment, not the thousands of a Monte Carlo!
(i.e. low bias at the price of high variance is bad)

Logic of The Test of Adding A Term

Details found in Simon Sirca's Probability for Physicist Book.

0) We have two models, a simple one with p-1 parameters, and a complex one with p parameters.

1) Null hypothesis (H0) = "simpler model (lower degree polynomial) is sufficient".

2) Compute the ratio F = [chi2(p-1) - chi2(p)] / chi2(p) * (N-p-1)

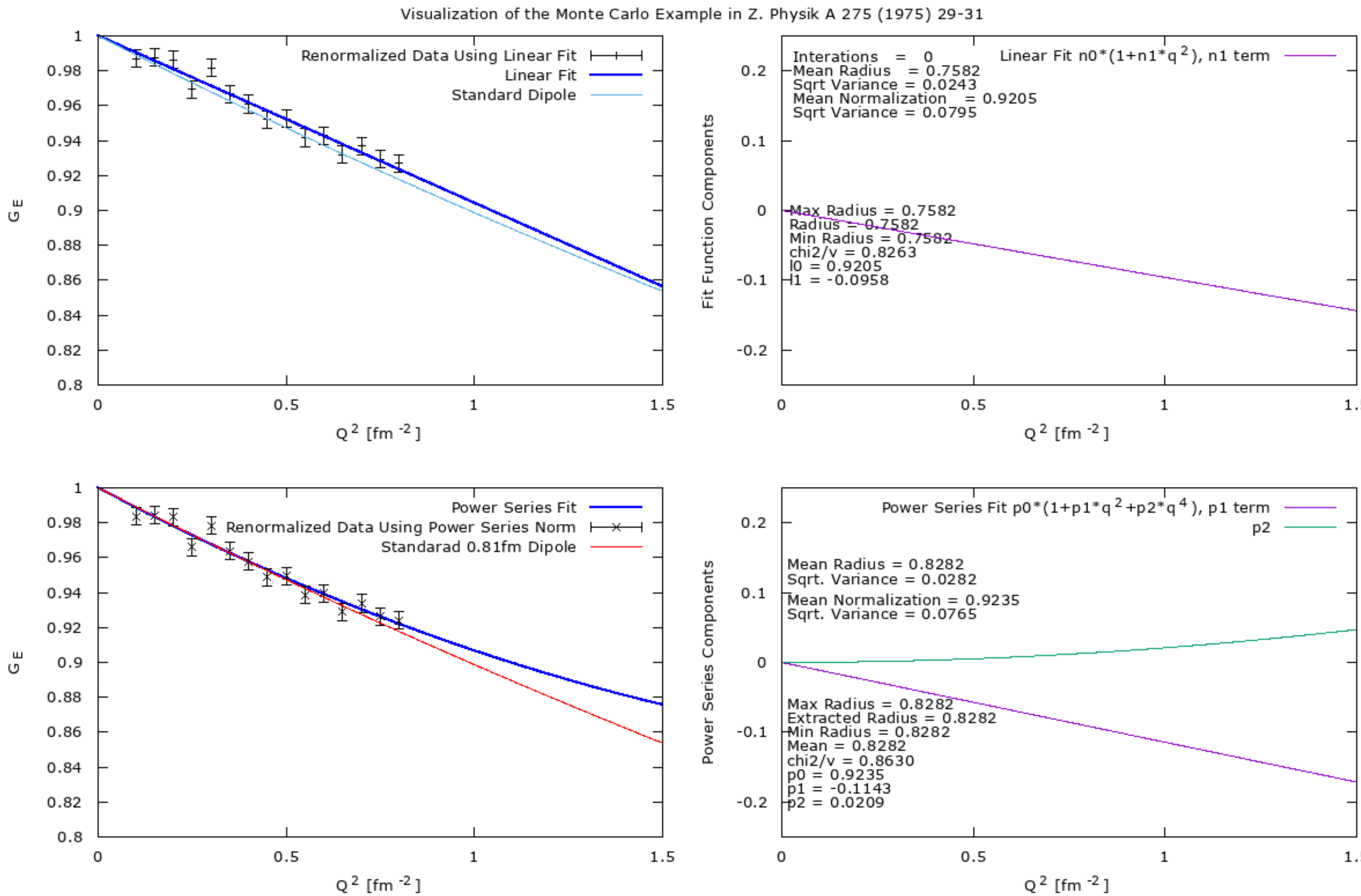
3) If H0 is correct, the ratio F will be distributed according to the F(1, N-p-1) distribution, thus:

4) We reject H0 (i.e. we likely need a more sophisticated model) if (F > F_{1-alpha}(1, N-p-1)) where CL = 1 - alpha, and typically alpha = 0.05, CL = 1-alpha = 0.95. The value of the cumulative distribution function F_{1-alpha}(1, N-p-1)) is calculated in:

Mathematica: Quantile[FRatioDistribution[1, N-p-1], 0.95]

R: qf(0.95, df1=1, df2=N-p-1)

The Bias Variance Trade-Off

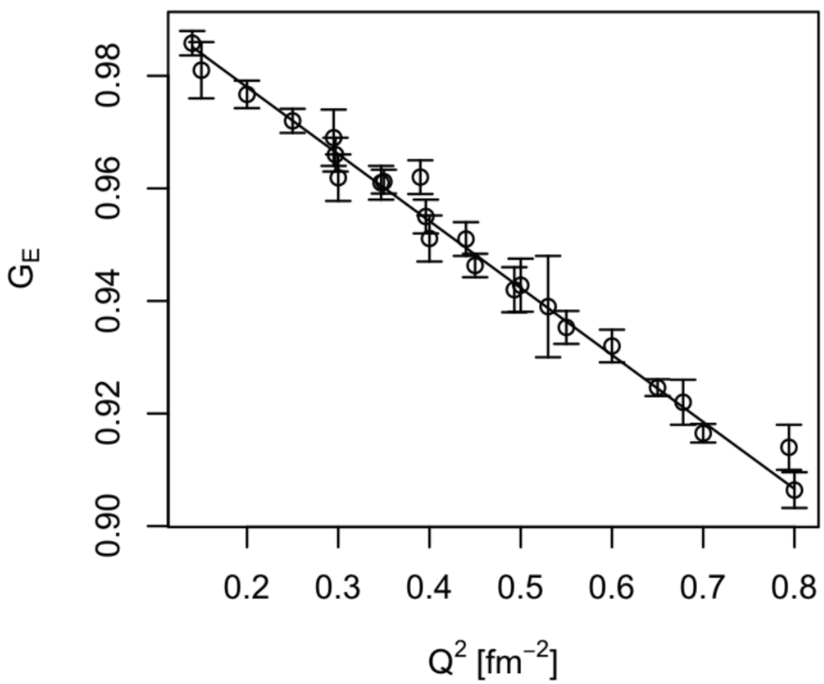


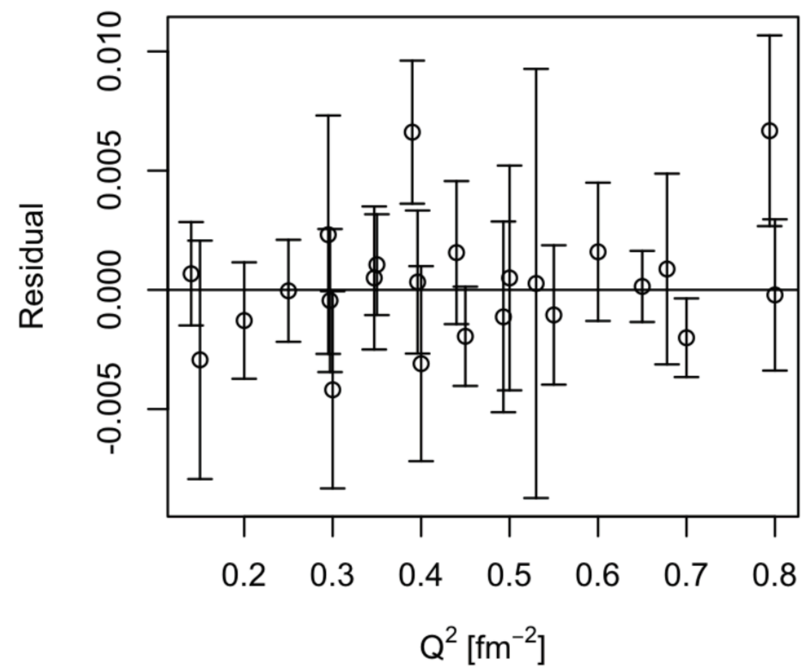
Real World Example

G. G. Simon, C. Schmitt, F. Borkowski, and V. H. Walther, Nucl. Phys. **A333** (1980) 381.
J. J. Murphy, Y. M. Shin, and D. M. Skopik, Phys. Rev. **C9** (1974) 2125.

$$f(Q^2) = n_0 G_E(Q^2) \approx n_0\left(1 + \sum_{i=1}^{m} a_i Q^{2i}\right)$$

N	j	χ^2	χ^2/ν	n_0	a_1	a_2
24	2	13.71	0.623	1.002(2)	−0.119(4)	
24	3	13.71	0.652	1.002(5)	−0.120(20)	0.00(2)



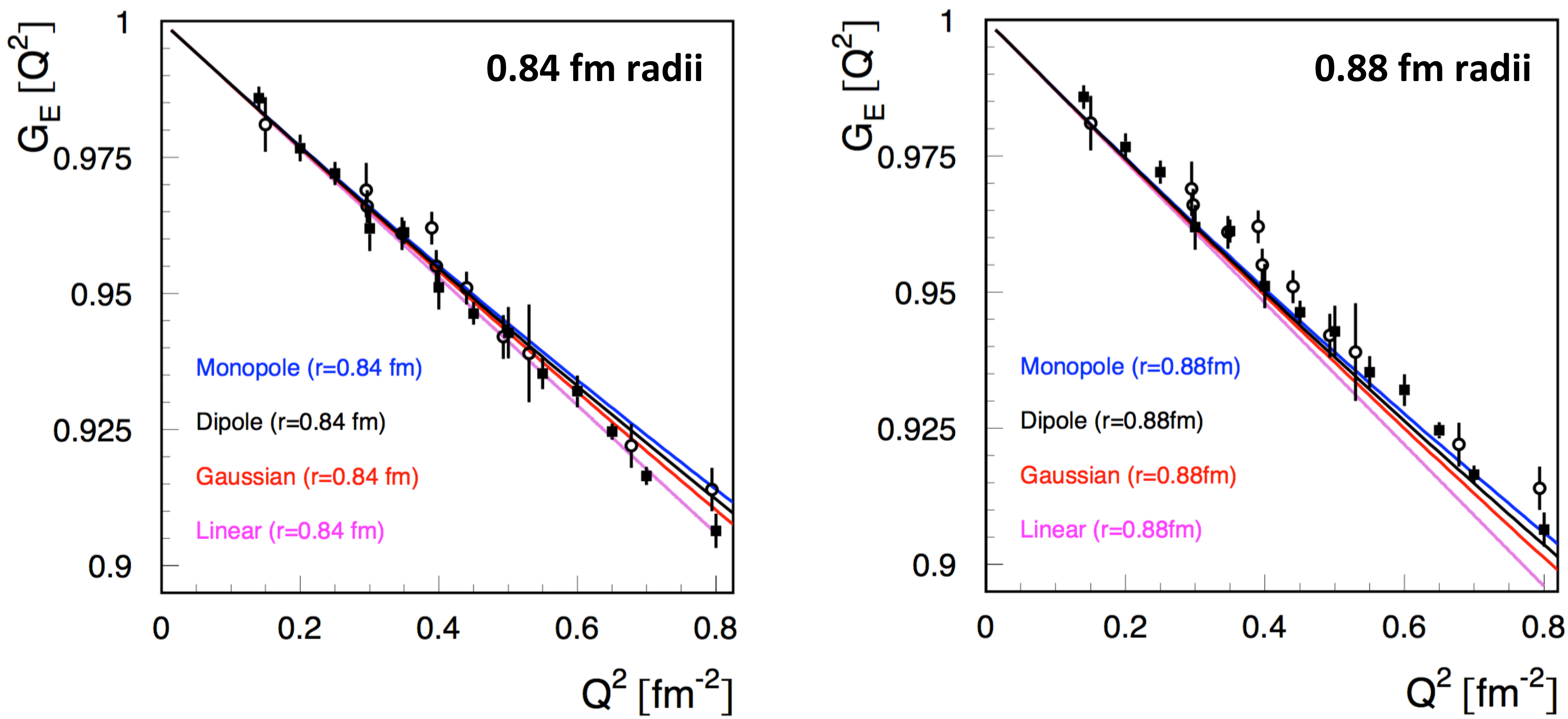


F-test rejects **fitting** with the more complex j=3 (j=m+1) function, that does NOT mean $a_2 = 0$.

Pohl *et.al's* 0.84 fm radius would predict an a_1 value of - 0.1176 since radius = sqrt(-6a_1)

Plotting Published Results & Standard Functions

- These are NOT regressions, just the data as published and standard curves. -



Closed Circles Mainz 1980 results and open circles Saskatoon 1974 results.

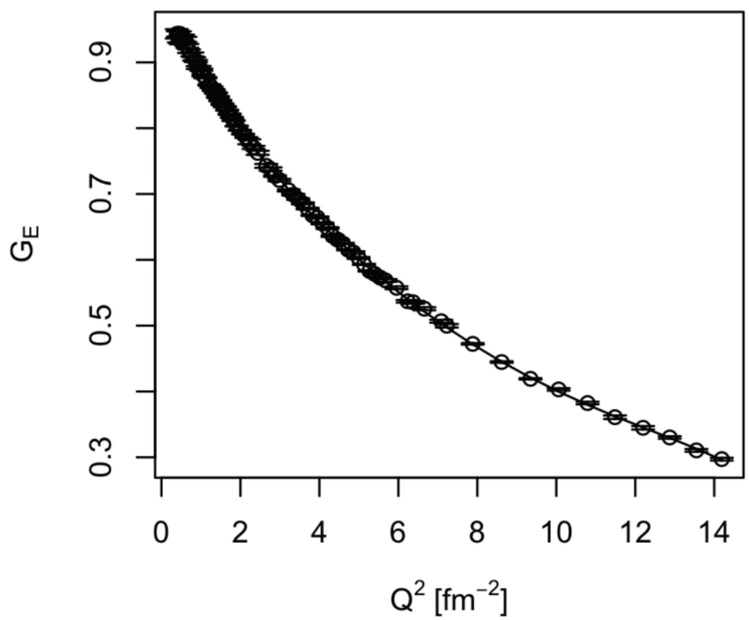
Note how for a fixed radius, all functions come together as Q^2 gets $< 0.4\ \mathrm{fm}^{-2}$.

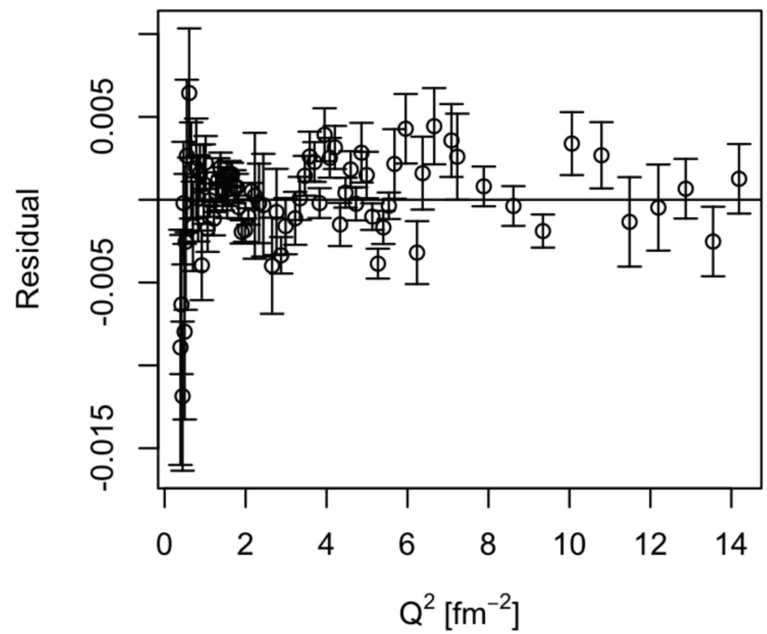
Mainz 2014 G_E Rosenbluth Data

Data found in J. Bernauer *et al.*, Phys Rev. **C90** (2014) 015206 supplemental material.

$$f(Q^2) = n_0 G_E(Q^2) \approx n_0 \left(1 + \sum_{i=1}^{m} a_i Q^{2i}\right)$$

Using F-test one rejects the 6th order polynomial, lower orders should be investigated!





N	j	χ^2	χ^2/ν	n_0	a_1	a_2	a_3	a_4	a_5	a_6
77	5	49.57	0.688	0.991(2)	$-0.113(1)$	$0.88(1) \cdot 10^{-2}$	$-0.44(2) \cdot 10^{-3}$	$9.7(8) \cdot 10^{-6}$		
77	6	41.34	0.582	0.996(2)	$-0.121(1)$	$1.25(1) \cdot 10^{-2}$	$-1.14(2) \cdot 10^{-3}$	$6.8(1) \cdot 10^{-5}$	$-1.62(7) \cdot 10^{-6}$	
77	7	41.32	0.590	0.995(3)	$-0.119(1)$	$1.18(1) \cdot 10^{-2}$	$-0.93(2) \cdot 10^{-3}$	$3.9(1) \cdot 10^{-5}$	$0.12(6) \cdot 10^{-6}$	$-4.2(5) \cdot 10^{-8}$

BUT one should be very wary of using a high order polynomials to extrapolate beyond the data.

Fixed Radius Fits

DWH *et al.*, Phys. Rev. C (2016).

- Again using the Mainz 2014 Rosenbluth results.
- Fit a power series with radius fixed to the two competing hypotheses
  - 0.84 fm from Muonic hydrogen
  - 0.88 fm from Atomic hydrogen

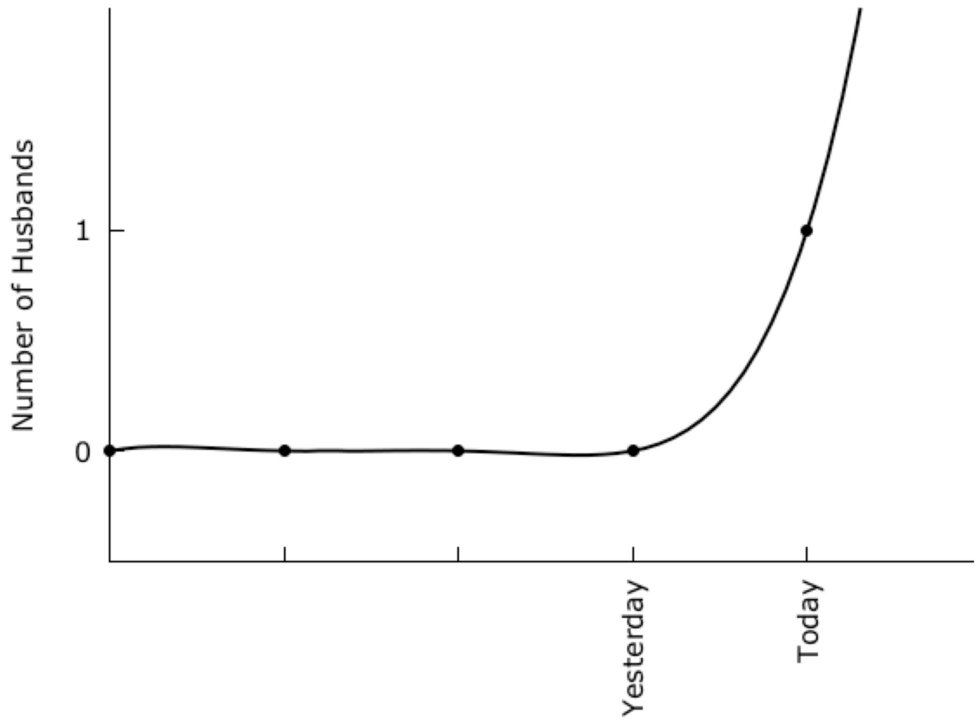
$$f(Q^2) = n_0 G_E(Q^2) \approx n_0 \left(1 + \sum_{i=1}^{m} a_i Q^{2i}\right)$$

Fixed Radius	χ^2	χ^2/ν	n_0	a_2	a_3	a_4	a_5
0.84 fm	56.34	0.783	0.994(1)	$1.12(1) \cdot 10^{-2}$	$-0.93(2) \cdot 10^{-3}$	$5.0(1) \cdot 10^{-5}$	$1.20(5) \cdot 10^{-6}$
0.88 fm	142.1	1.97	1.003(1)	$1.62(1) \cdot 10^{-2}$	$-1.78(1) \cdot 10^{-3}$	$1.14(1) \cdot 10^{-4}$	$-2.90(7) \cdot 10^{-6}$

But high order polynomials cannot extrapolate.

Minimum Chi^2 Minimum = ZERO

Note: One can often find a minimum in reduced chi^2.



But this one is even better then my two parameter exponential!

Fit function f(x) = f0 + f1*x + f2*x^2 + f3*x^3 + f3*x^4

WARNING: Chi^2 tests will not save you from using an in-appropriate function!

Padé Approximant & Continued Fractions

Pade' Approximant

When it exists, the Pade' approximant (N,M) of a Tayler series is unique.

$$f(x) = \frac{a_0 + a_1 x^1 + a_2 x^2 \ldots + a^M * x^M}{1 + b_1 x^1 + b_2 x^2 \ldots + b^N * x^N}$$

In our case we want $f(x) = n_0 G_E(Q^2)$, so

$$f(x) = n_0 \frac{1 + a_1 Q_2 + a_2 Q^4 \ldots + a^{M*2} * Q^{M*2}}{1 + b_1 Q_2 + b_2 Q^4 \ldots + b^{N*2} * x^{N*2}}$$

(Henri Padé ~ 1860)

Continued Fraction

$$f(Q^2) = \frac{c_1}{1 + \frac{c_2 Q^2}{1+\frac{c_3 Q^2}{1+\frac{c_4 Q^2}{1+\ldots}}}}$$

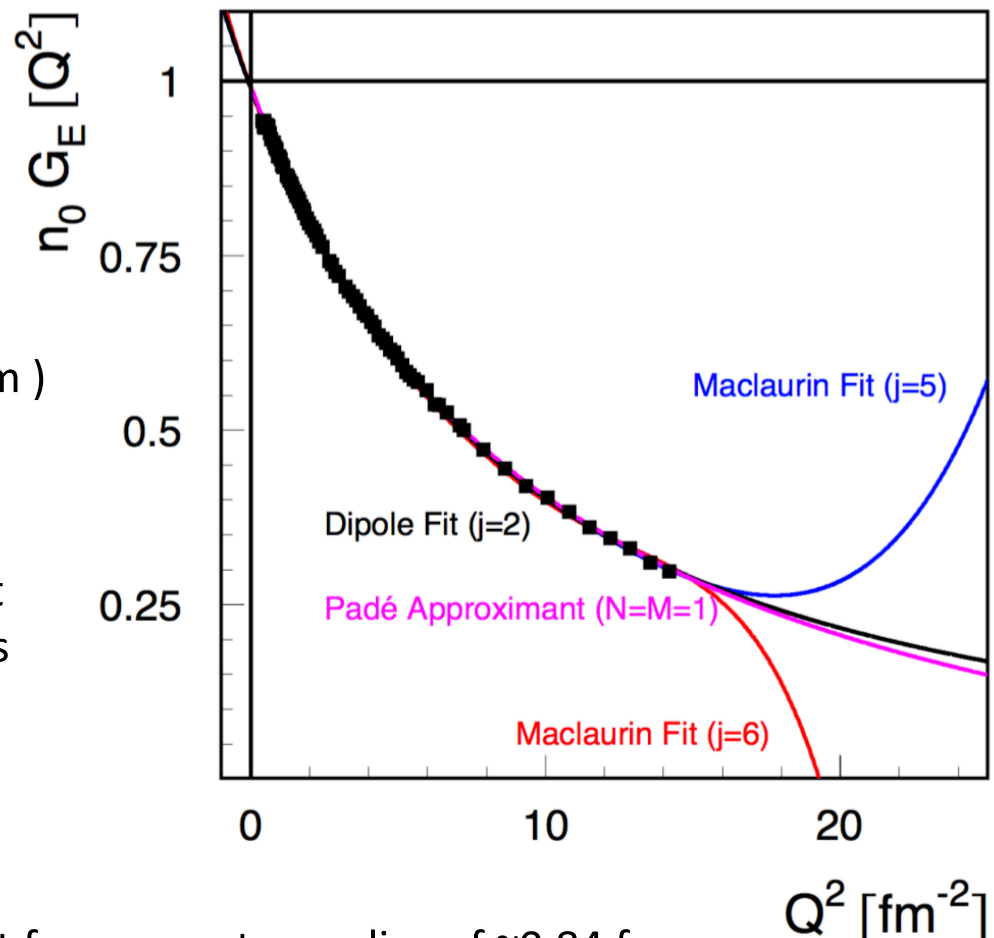
(Ancient Greeks)

Further reading: **Extrapolation algorithms and Padé approximations: a historical survey**
C. Brezinski, Applied Numerical Mathematics 20 (1996) 299.

Maclaurin, Padé Approximant & Dipole Fits

Using the Mainz14 Rosenluth Results (**where G_E & G_M well constrained by the data**).

$$f(Q^2) = n_0 G_E(Q^2) \approx n_0 \left(1 + \sum_{i=1}^{m} a_i Q^{2i}\right)$$



Used f test to rule out j=7 (m = 6 & n_0 term)

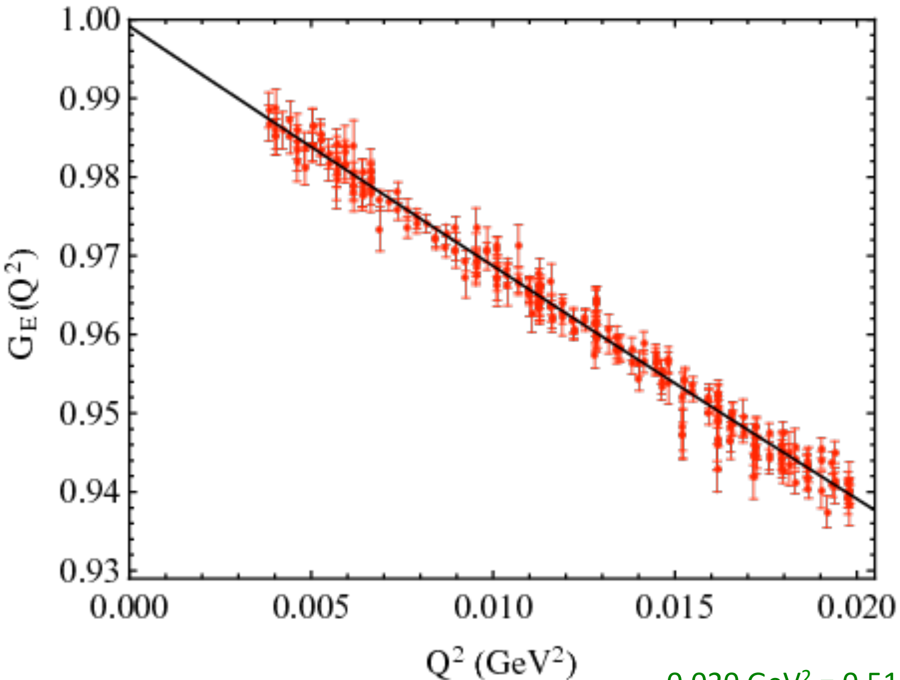
WARNING: F test can reject functions, but It doesn't tell you which of the remaining is "best" or most appropriate.

(i.e. inspect the results!)

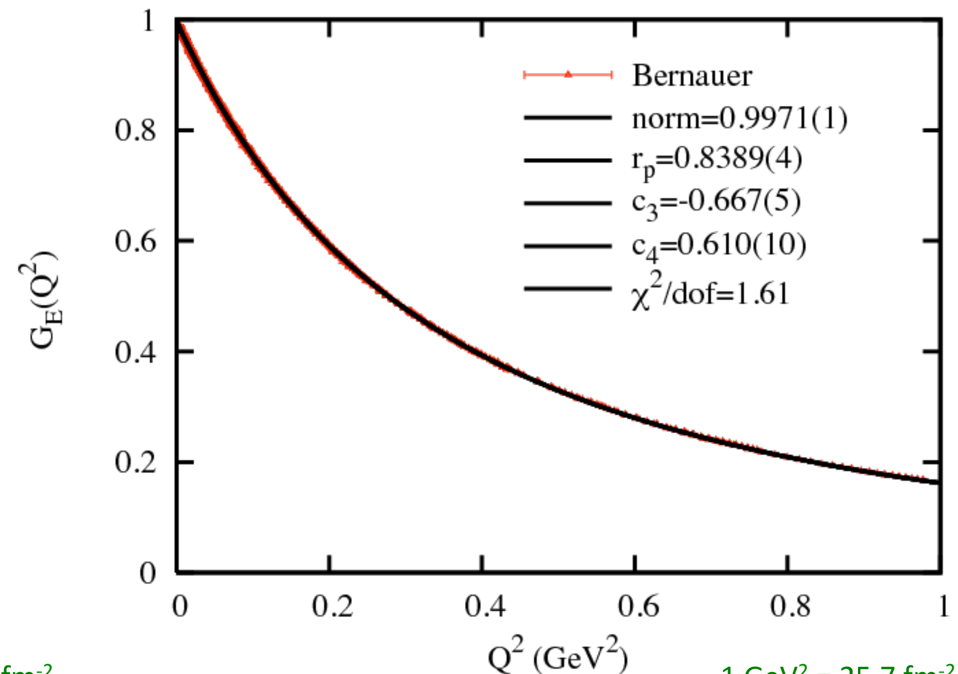
These fits all give results that favor a proton radius of ~0.84 fm.
Note how Padé and dipole fits extrapolate nicely, while the Maclaurin quickly diverge.

William & Mary Analysis

K. Griffioen, C. Carlson, S. Maddox, Phys. Rev. **C93** (2016) 065207.



0.020 GeV^2 = 0.51 fm^{-2}



1 GeV^2 = 25.7 fm^{-2}

Linear and Quadratic Fits of Low Q^2 Data & Continued Fraction Fits To Q^2 of 1 GeV^2 (25.7 fm^{-2})

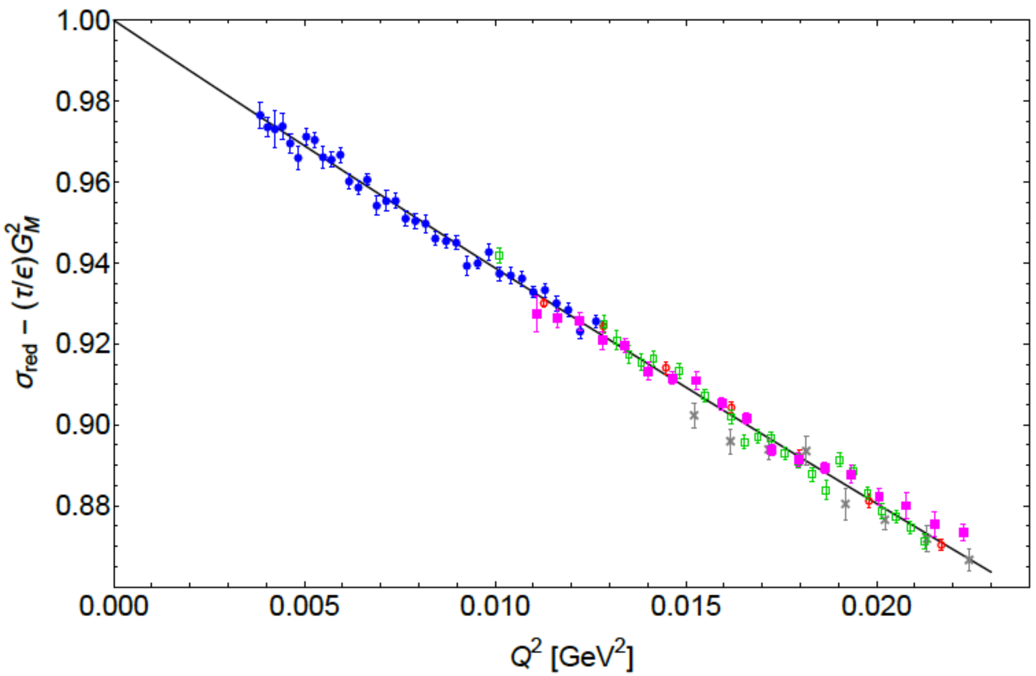
All three results consistent with the 0.84 fm radius of the Muonic hydrogen Lamb shift.

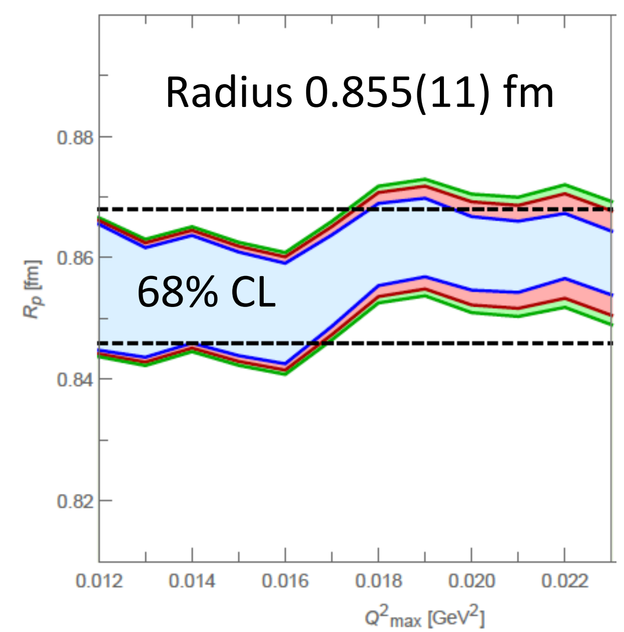
NOTE: Publishing simple explanations that disagreed with complex 0.88 fm results proved to be amazingly challenging.

ChPT Inspired Analysis

M. Horbatsch, E. A. Hessels, and A. Pineda, Phys. Rev. C 95 (2017) 035203.

- Idea to use ChPT to constrain moments.
- Non-linear model mathematics fits in order to float normalizations.
- Colors Indicate Different Floating Normalizations (as defined by Mainz)
- Not clear yet to me that the uncertainty is really Gaussian about the mean.





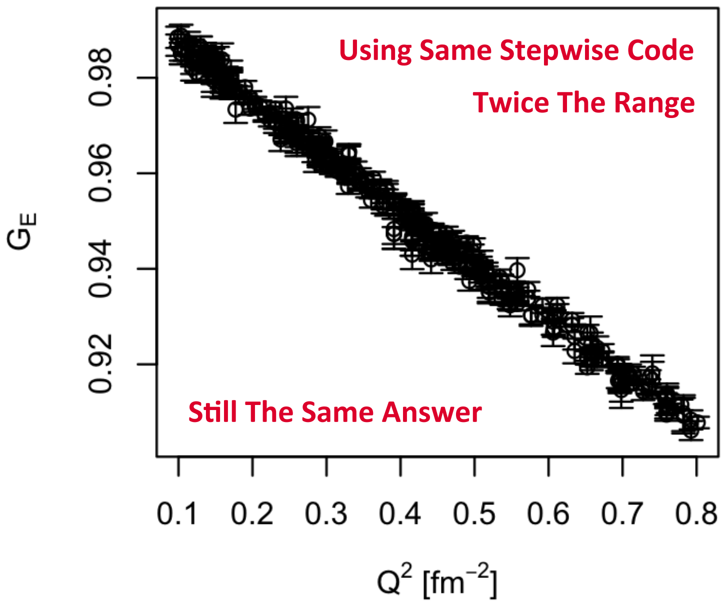
Beyond Simple Fitting: Stepwise Regression

2015

Language Rank		Types	Spectrum Ranking
1.	Java		100.0
2.	C		99.9
3.	C++		99.4
4.	Python		96.5
5.	C#		91.3
6.	R		84.8
7.	PHP		84.5
8.	JavaScript		83.0
9.	Ruby		76.2
10.	Matlab		72.4

IEEE Rankings are based mostly on CPU usage (i.e. big data)

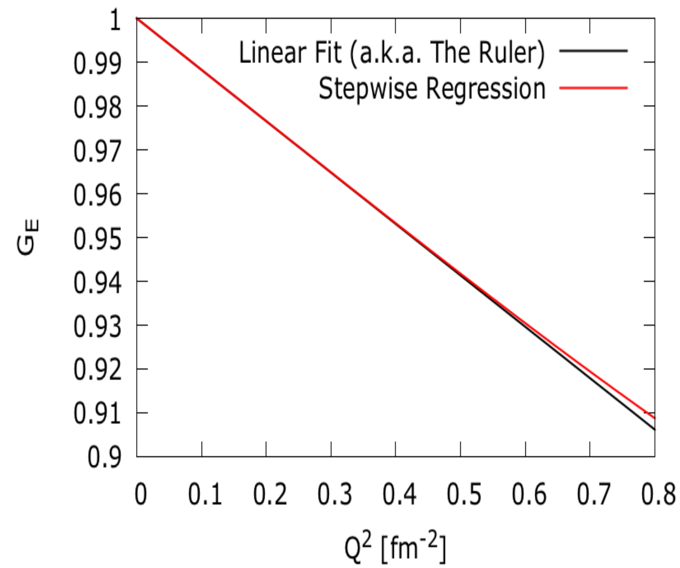
Stepwise Regression of G_E from 2014 Data

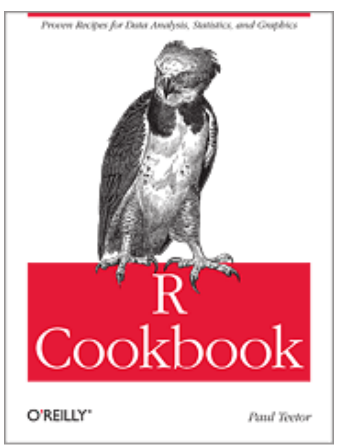



```
Start:  AIC=36.77
data$y ~ data$x

              Df Sum of Sq    RSS    AIC
+ I(data$x^4)   1   10.3725 358.06 29.236
+ I(data$x^3)   1   10.2911 358.14 29.312
+ I(data$x^5)   1   10.2718 358.16 29.330
+ I(data$x^6)   1   10.0519 358.38 29.535
+ I(data$x^2)   1    9.9568 358.48 29.624
+ I(data$x^7)   1    9.7627 358.67 29.804
+ I(data$x^8)   1    9.4401 359.00 30.105
+ I(data$x^9)   1    9.1075 359.33 30.414
+ I(data$x^10)  1    8.7790 359.66 30.719
+ I(data$x^11)  1    8.4620 359.97 31.013
<none>                      368.44 36.774

Step:  AIC=29.24
data$y ~ data$x + I(data$x^4)

              Df Sum of Sq    RSS    AIC
<none>                      358.06 29.236
+ I(data$x^2)   1 0.0088531 358.05 31.228
+ I(data$x^3)   1 0.0028516 358.06 31.233
+ I(data$x^11)  1 0.0007801 358.06 31.235
+ I(data$x^5)   1 0.0006383 358.06 31.236
+ I(data$x^6)   1 0.0004668 358.06 31.236
+ I(data$x^7)   1 0.0003015 358.06 31.236
+ I(data$x^10)  1 0.0001705 358.06 31.236
+ I(data$x^8)   1 0.0001061 358.06 31.236
+ I(data$x^9)   1 0.0000000 358.06 31.236
```

Akaike Information Criterion Selected Model

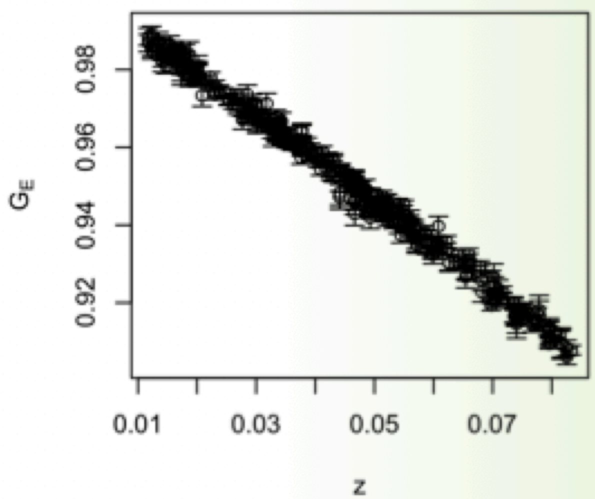
```
Call:
lm(formula = data$y ~ data$x + I(data$x^4), weights = 1/data$dy^2)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-3.02110 -0.73469 -0.08639  0.66588  3.08298

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.9988419  0.0003534 2826.253  < 2e-16 ***
data$x      -0.1172672  0.0010936 -107.229  < 2e-16 ***
I(data$x^4)  0.0063583  0.0020534    3.097  0.00213 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.04 on 331 degrees of freedom
Multiple R-squared:  0.9932,    Adjusted R-squared:  0.9932
F-statistic: 2.434e+04 on 2 and 331 DF,  p-value: < 2.2e-16
```

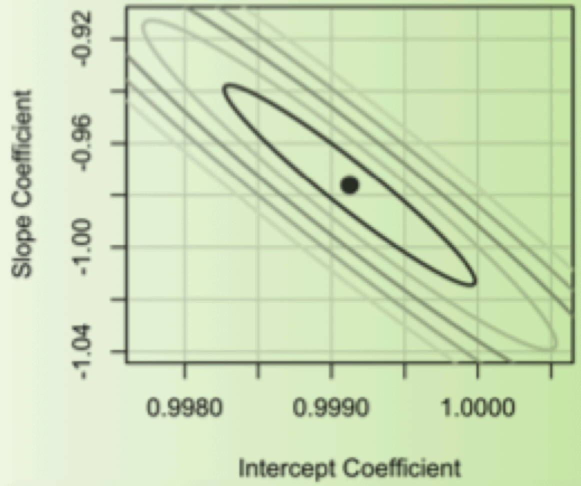




Pohl *et.al's* 0.84 fm radius would predict a slope of - 0.1176 !!

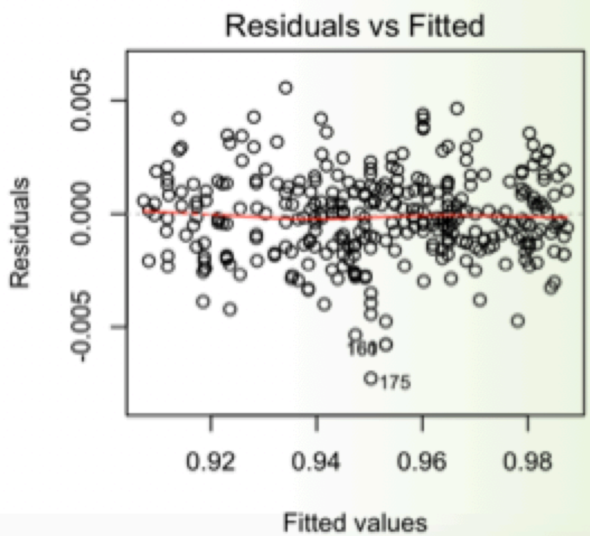
Conformal Mapping vs. Stepwise Regression

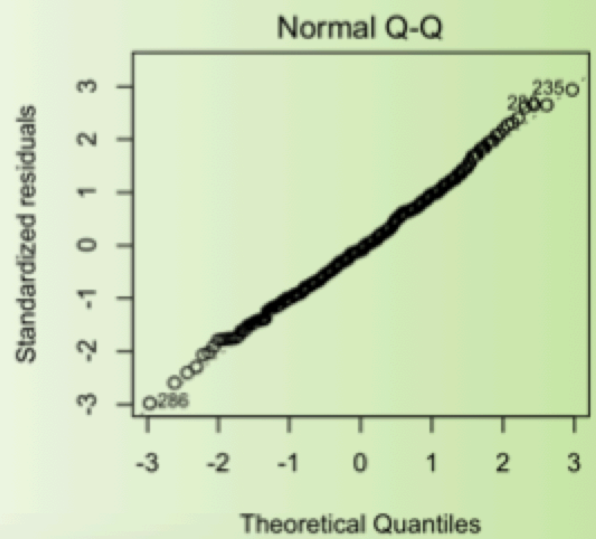


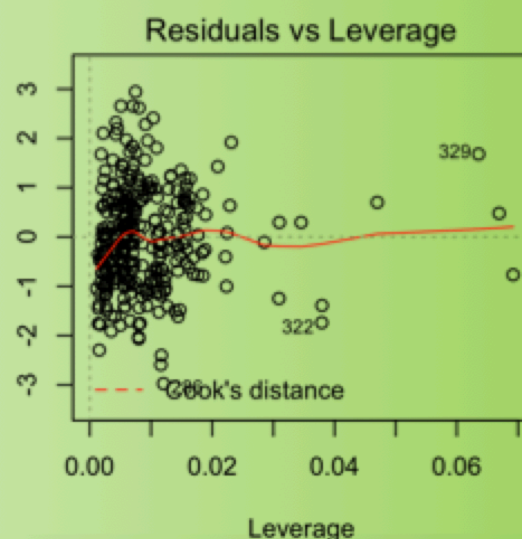



```
Call:
lm(formula = data$y ~ data$x + I(data$x^2), weights = 1/data$dy^2)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-3.07679 -0.75645 -0.09102  0.67025  3.05076

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.999125   0.000565 1768.416  < 2e-16 ***
data$x      -0.976097   0.025338  -38.523  < 2e-16 ***
I(data$x^2) -1.520998   0.260916   -5.829 1.32e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.041 on 331 degrees of freedom
Multiple R-squared:  0.9932,Adjusted R-squared:  0.9932
F-statistic: 2.431e+04 on 2 and 331 DF,  p-value: < 2.2e-16
```

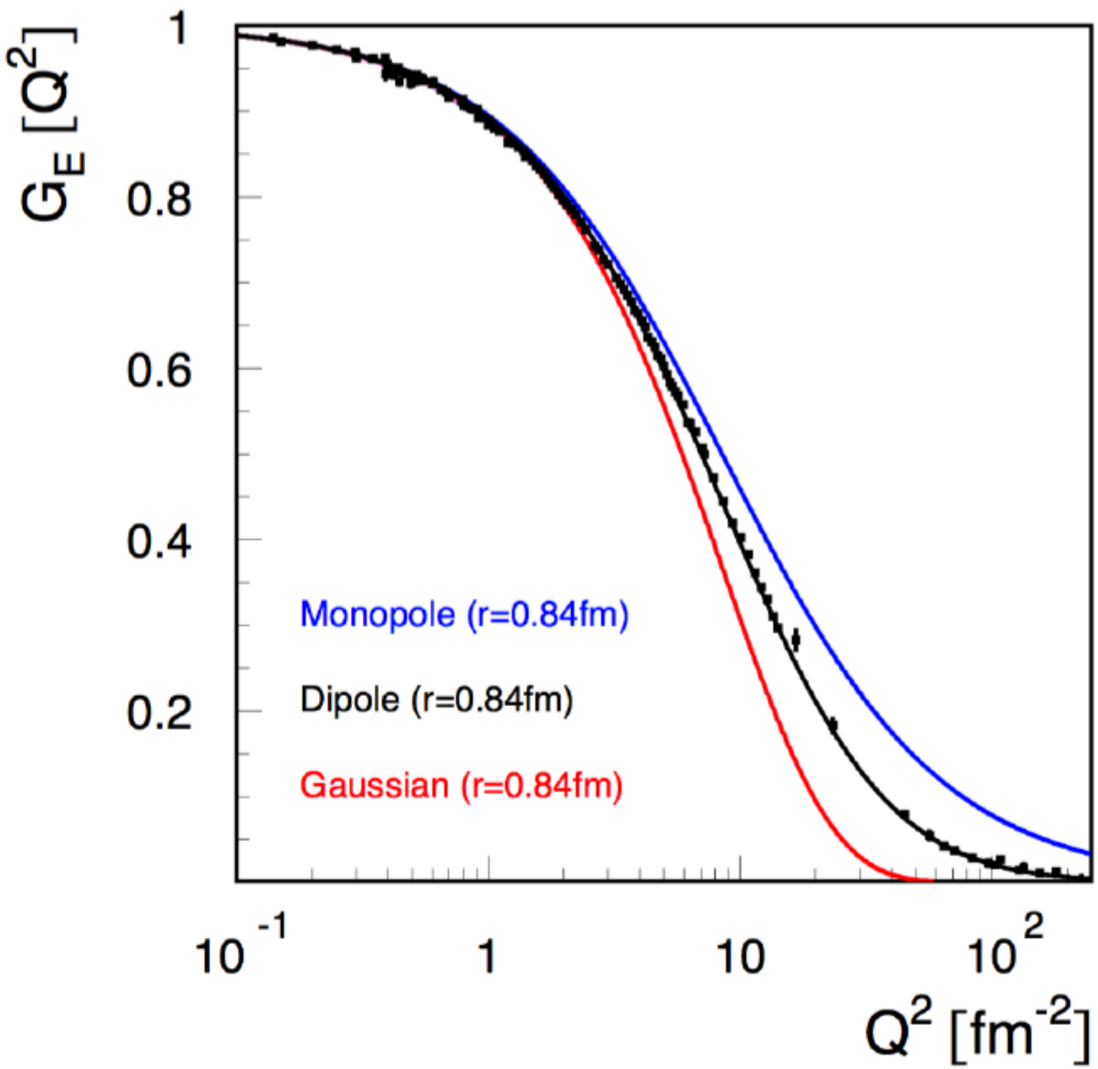
Beautiful Fit with my same Stepwise regression code and I get a radius of 0.84 fm







The "Textbook" Plot

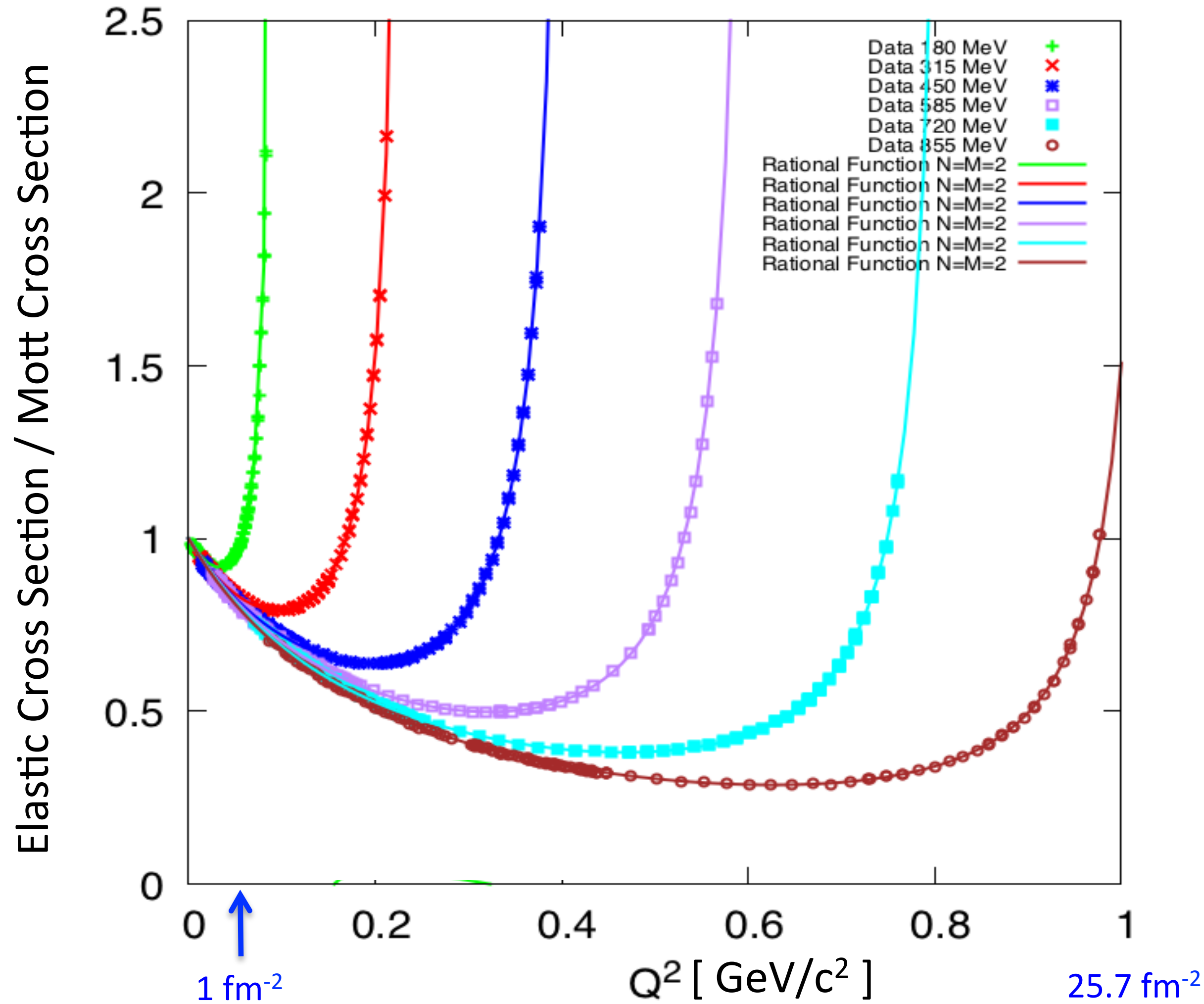


But these values are too small!

- Using standard dipole (0.81 fm radius) it was *shown* linear fits (0.84 fm) are biased since, **"The commonly used polynomial of first order yields radius values which are too small."** – Z. Phys. A 275 (1975) 29.
- **How did they know it was *too small* ?!** (i.e. you need to know the true radius to know what is truly too small.)
- Monte Carlo's are still being used to do this *proof* except the functions are often even more complex and tend to bend near the end of the data.
- Using an F-Test or stepwise regression, the classic data could not support the extraction of a quadratic term for < 0.8 fm^{-2} data.
- Low Q^2 G_E fits of Griffioen, Carlson, Maddox also do not show a need to include a quadratic term with their $Q^2 < 0.5\ fm^{-2}$ fits.
- New paper using ChPT calculated moments indicate the expectation is that the moments are relatively small.
- **So why did Mainz report a large radius with their new data?!**

Mainz 2014 1422 Data Points Plotted vs. Q^2

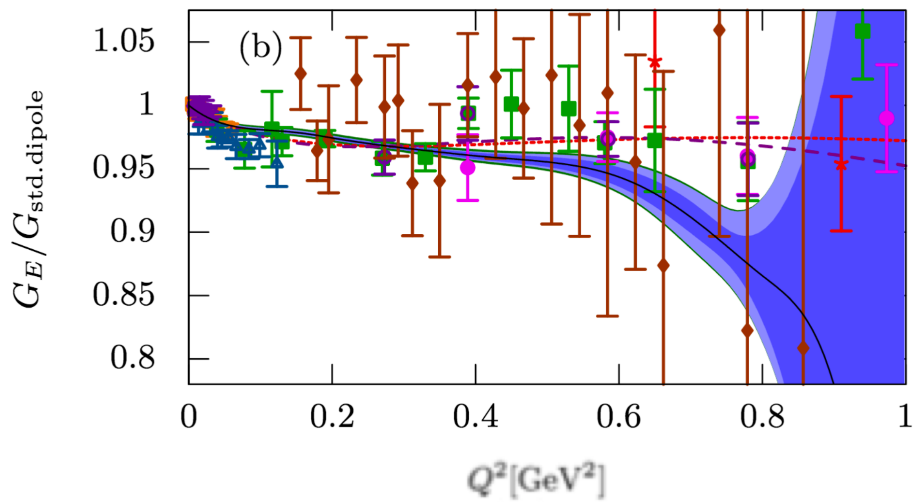
Absolutely beautiful data!

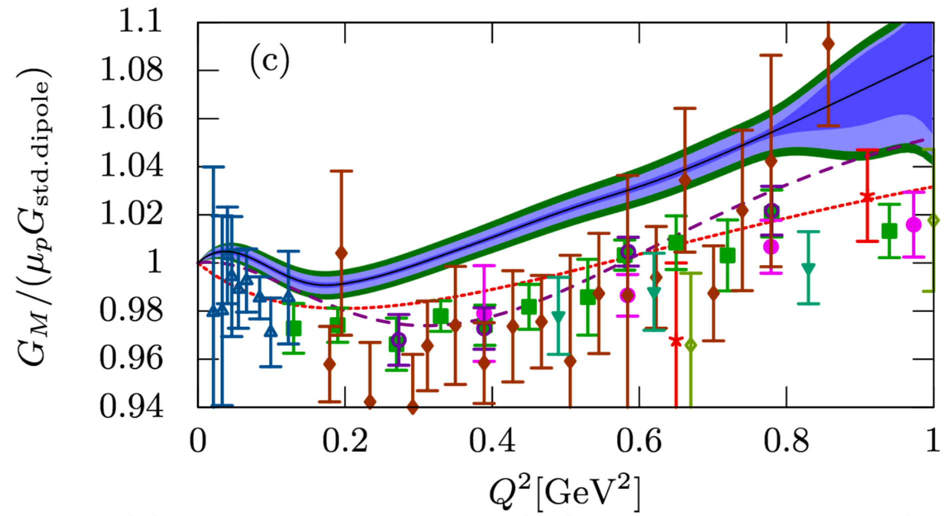


NOTE: $Q^2 = Q^2(E,\text{theta})$ has a kinematic max., $Q^2_{max}(E,180^o)$, which these Padé fits nicely reproduce.

Mainz 2014 G_E & G_M (Blue Band)

Example Results from Bernauer *et al.*, Phys Rev. C90 (2014) 015206.





Blue Bands are the Mainz results and the points are world data.

Mainz data is available within the PRC supplemental material:
http://journals.aps.org/prc/supplemental/10.1103/PhysRevC.90.015206
And a re-binned version is available from Lee, Arrington, and Hill:
http://journals.aps.org/prd/supplemental/10.1103/PhysRevD.92.013013

Bernauer *et al.*'s Large Proton Radius Comes From A 51 Parameter Non-Linear Regression!

$$G_E[Q^2] = 1+\sum_{i=1}^{10} e_i (Q^2)^i \qquad G_M[Q^2] = \mu_0(1+\sum_{i=1}^{10} m_i (Q^2)^i$$

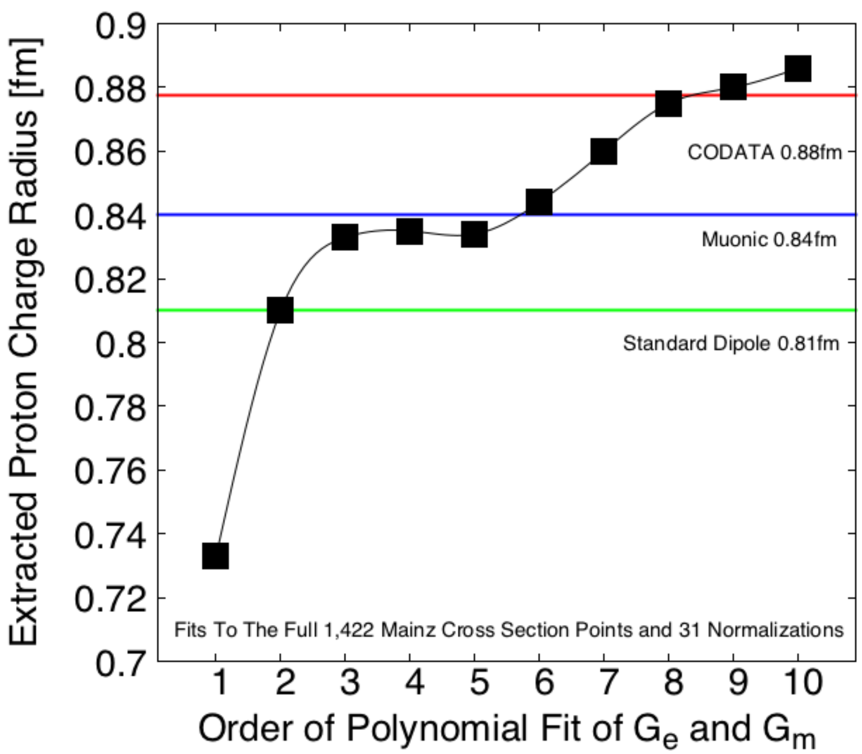
plus 31 floating normalizations

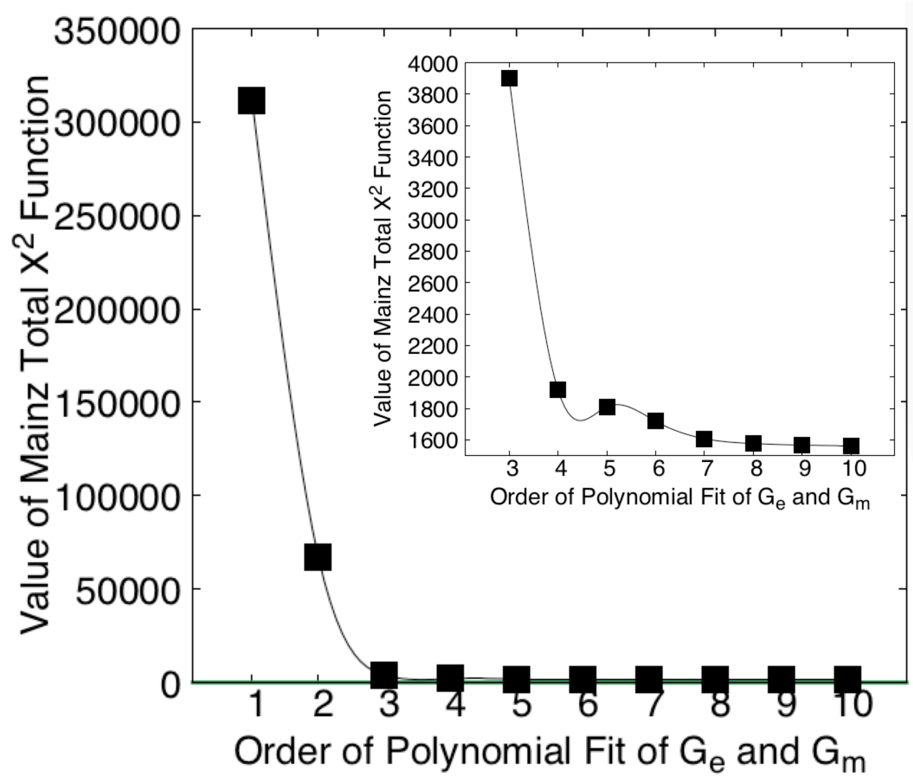
$$
\begin{aligned}
\left(\frac{d\sigma}{d\Omega}\right)/\left(\frac{d\sigma}{d\Omega}\right)_{\rm Mott}\cdot\frac{E}{E'} &= \left[\frac{G_E^2+\tau G_M^2}{1+\tau}+2\tau G_M^2\tan^2(\theta/2)\right]\\
&= 1/(1+\tau)+(\mu_0^2\tau)/(1+\tau)+2\mu_0^2\tau\tan^2(\theta/2)\\
&\quad +Q^2[(2e_1)/(1+\tau)+(2m_1\mu_0^2\tau)/(1+\tau)+4m_1\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^4[({e_1}^2)/(1+\tau)+(2e_2)/(1+\tau)+({m_1}^2\mu_0^2\tau)/(1+\tau)+(2m_2\mu_0^2\tau)/((1+\tau))+2{m_1}^2\mu_0^2\tau\tan^2(\theta/2)+4m_2\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^6[(2e_1e_2)/(1+\tau)+(2e_3)/(1+\tau)+(2m_1m_2\mu_0^2\tau)/(1+\tau)+(2m_3\mu_0^2\tau)/((1+\tau))+4m_1m_2\mu_0^2\tau\tan^2(\theta/2)+4m_3\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^8[({e_2}^2)/(1+\tau)+(2e_1e_3)/(1+\tau)+(2e_4)/(1+\tau)+({m_2}^2\mu_0^2\tau)/(1+\tau)+(2m_1m_3\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_4\mu_0^2\tau)/((1+\tau))+2{m_2}^2\mu_0^2\tau\tan^2(\theta/2)+4m_1m_3\mu_0^2\tau\tan^2(\theta/2)+4m_4\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{10}[(2e_2e_3)/(1+\tau)+(2e_1e_4)/(1+\tau)+(2e_5)/(1+\tau)+(2m_2m_3\mu_0^2\tau)/(1+\tau)+2m_1m_4\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_5\mu_0^2\tau)/((1+\tau))+4m_2m_3\mu_0^2\tau\tan^2(\theta/2)+4m_1m_4\mu_0^2\tau\tan^2(\theta/2)+4m_5\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{12}[({e_3}^2)/(1+\tau)+(2e_2e_4)/(1+\tau)+(2e_1e_5)/(1+\tau)+(2e_6)/(1+\tau)+({m_3}^2\mu_0^2\tau)/(1+\tau)+(2m_2m_4\mu_0^2\tau)/((1+\tau))\\
&\qquad +(2m_1m_5\mu_0^2\tau)/(1+\tau)+(2m_6\mu_0^2\tau)/(1+\tau)+2{m_3}^2\mu_0^2\tau\tan^2(\theta/2)+4m_2m_4\mu_0^2\tau\tan^2(\theta/2)+4m_1m_5\mu_0^2\tau\tan^2(\theta/2)+4m_6\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{14}[(2e_3e_4)/(1+\tau)+(2e_2e_5)/(1+\tau)+(2e_1e_6)/(1+\tau)+(2e_7)/(1+\tau)+(2m_3m_4\mu_0^2\tau)/(1+\tau)+(2m_2m_5\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_1m_6\mu_0^2\tau)/(1+\tau)+(2m_7\mu_0^2\tau)/(1+\tau)+4m_3m_4\mu_0^2\tau\tan^2(\theta/2)+4m_2m_5\mu_0^2\tau\tan^2(\theta/2)+4m_1m_6\mu_0^2\tau\tan^2(\theta/2)+4m_7\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{16}[({e_4}^2)/(1+\tau)+(2e_3e_5)/(1+\tau)+(2e_2e_6)/(1+\tau)+(2e_1e_7)/(1+\tau)+(2e_8)/(1+\tau)+({m_4}^2\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_3m_5\mu_0^2\tau)/(1+\tau)+(2m_2m_6\mu_0^2\tau)/((1+\tau))+(2m_1m_7\mu_0^2\tau)/(1+\tau)+(2m_8\mu_0^2\tau)/(1+\tau)+2{m_4}^2\mu_0^2\tau\tan^2(\theta/2)+4m_3m_5\mu_0^2\tau\tan^2(\theta/2)\\
&\qquad +4m_2m_6\mu_0^2\tau\tan^2(\theta/2)+4m_1m_7\mu_0^2\tau\tan^2(\theta/2)+4m_8\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{18}[(2e_4e_5)/(1+\tau)+(2e_3e_6)/(1+\tau)+(2e_2e_7)/(1+\tau)+(2e_1e_8)/(1+\tau)+(2e_9)/(1+\tau)+(2m_4m_5\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_3m_6\mu_0^2\tau)/(1+\tau)+(2m_2m_7\mu_0^2\tau)/((1+\tau))+(2m_1m_8\mu_0^2\tau)/(1+\tau)+(2m_9\mu_0^2\tau)/(1+\tau)+4m_4m_5\mu_0^2\tau\tan^2(\theta/2)\\
&\qquad +4m_3m_6\mu_0^2\tau\tan^2(\theta/2)+4m_2m_7\mu_0^2\tau\tan^2(\theta/2)+4m_1m_8\mu_0^2\tau\tan^2(\theta/2)+4m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{20}[(2e_{10})/(1+\tau)+({e_5}^2)/(1+\tau)+(2e_4e_6)/(1+\tau)+(2e_3e_7)/(1+\tau)+(2e_2e_8)/(1+\tau)+(2e_1e_9)/(1+\tau)\\
&\qquad +(2m_{10}\mu_0^2\tau)/(1+\tau)+({m_5}^2\mu_0^2\tau)/((1+\tau))+(2m_4m_6\mu_0^2\tau)/(1+\tau)+(2m_3m_7\mu_0^2\tau)/(1+\tau)+(2m_2m_8\mu_0^2\tau)/((1+\tau))\\
&\qquad +(2m_1m_9\mu_0^2\tau)/(1+\tau)4m_{10}\mu_0^2\tau\tan^2(\theta/2)+2{m_5}^2\mu_0^2\tau\tan^2(\theta/2)+4m_4m_6\mu_0^2\tau\tan^2(\theta/2)+4m_3m_7\mu_0^2\tau\tan^2(\theta/2)+4m_2m_8\mu_0^2\tau\tan^2(\theta/2)+4m_1m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{22}[(2e_1e_{10})/(1+\tau)+(2e_5e_6)/(1+\tau)+(2e_4e_7)/(1+\tau)+(2e_3e_8)/(1+\tau)+(2e_2e_9)/(1+\tau)+(2m_1m_{10}\mu_0^2\tau)/((1+\tau))\\
&\qquad +(2m_5m_6\mu_0^2\tau)/(1+\tau)+(2m_4m_7\mu_0^2\tau)/(1+\tau)+(2m_3m_8\mu_0^2\tau)/((1+\tau))+(2m_2m_9\mu_0^2\tau)/(1+\tau)4m_1m_{10}\mu_0^2\tau\tan^2(\theta/2)+\\
&\qquad +4m_5m_6\mu_0^2\tau\tan^2(\theta/2)+4m_4m_7\mu_0^2\tau\tan^2(\theta/2)+4m_3m_8\mu_0^2\tau\tan^2(\theta/2)+4m_2m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{24}[(2e_{10}e_2)/(1+\tau)+({e_6}^2)/(1+\tau)+(2e_5e_7)/(1+\tau)+(2e_4e_8)/(1+\tau)+(2e_3e_9)/(1+\tau)+(2m_{10}m_2\mu_0^2\tau)/((1+\tau))\\
&\qquad +({m_6}^2\mu_0^2\tau)/(1+\tau)+(2m_5m_7\mu_0^2\tau)/(1+\tau)+(2m_4m_8\mu_0^2\tau)/((1+\tau))+(2m_3m_9\mu_0^2\tau)/(1+\tau)+4m_{10}m_2\mu_0^2\tau\tan^2(\theta/2)\\
&\qquad +2{m_6}^2\mu_0^2\tau\tan^2(\theta/2)+4m_5m_7\mu_0^2\tau\tan^2(\theta/2)+4m_4m_8\mu_0^2\tau\tan^2(\theta/2)+4m_3m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{26}[(2e_{10}e_3)/(1+\tau)+(2e_6e_7)/(1+\tau)+(2e_5e_8)/(1+\tau)+(2e_4e_9)/(1+\tau)+(2m_{10}m_3\mu_0^2\tau)/(1+\tau)+(2m_6m_7\mu_0^2\tau)/((1+\tau))\\
&\qquad +(2m_5m_8\mu_0^2\tau)/(1+\tau)+(2m_4m_9\mu_0^2\tau)/(1+\tau)+4m_{10}m_3\mu_0^2\tau\tan^2(\theta/2)+4m_6m_7\mu_0^2\tau\tan^2(\theta/2)+4m_5m_8\mu_0^2\tau\tan^2(\theta/2)+4m_4m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{28}[(2e_{10}e_4)/(1+\tau)+({e_7}^2)/(1+\tau)+(2e_6e_8)/(1+\tau)+(2e_5e_9)/(1+\tau)+(2m_{10}m_4\mu_0^2\tau)/(1+\tau)+({m_7}^2\mu_0^2\tau)/((1+\tau))\\
&\qquad +(2m_6m_8\mu_0^2\tau)/(1+\tau)+(2m_5m_9\mu_0^2\tau)/(1+\tau)+4m_{10}m_4\mu_0^2\tau\tan^2(\theta/2)+2{m_7}^2\mu_0^2\tau\tan^2(\theta/2)+4m_6m_8\mu_0^2\tau\tan^2(\theta/2)+4m_5m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{30}[(2e_{10}e_5)/(1+\tau)+(2e_7e_8)/(1+\tau)+(2e_6e_9)/(1+\tau)+(2m_{10}m_5\mu_0^2\tau)/((1+\tau))+(2m_7m_8\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_6m_9\mu_0^2\tau)/(1+\tau)+4m_{10}m_5\mu_0^2\tau\tan^2(\theta/2)+4m_7m_8\mu_0^2\tau\tan^2(\theta/2)+4m_6m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{32}[(2e_{10}e_6)/(1+\tau)+({e_8}^2)/(1+\tau)+(2e_7e_9)/(1+\tau)+(2m_{10}m_6\mu_0^2\tau)/((1+\tau))+({m_8}^2\mu_0^2\tau)/(1+\tau)\\
&\qquad +(2m_7m_9\mu_0^2\tau)/(1+\tau)+4m_{10}m_6\mu_0^2\tau\tan^2(\theta/2)+2{m_8}^2\mu_0^2\tau\tan^2(\theta/2)+4m_7m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{34}[(2e_{10}e_7)/(1+\tau)+(2e_8e_9)/(1+\tau)+(2m_{10}m_7\mu_0^2\tau)/(1+\tau)+(2m_8m_9\mu_0^2\tau)/((1+\tau))+4m_{10}m_7\mu_0^2\tau\tan^2(\theta/2)+4m_8m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{36}[(2e_{10}e_8)/(1+\tau)+({e_9}^2)/(1+\tau)+(2m_{10}m_8\mu_0^2\tau)/(1+\tau)+({m_9}^2\mu_0^2\tau)/((1+\tau))+4m_{10}m_8\mu_0^2\tau\tan^2(\theta/2)+2{m_9}^2\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{38}[(2e_{10}e_9)/(1+\tau)+(2m_{10}m_9\mu_0^2\tau)/((1+\tau))+4m_{10}m_9\mu_0^2\tau\tan^2(\theta/2)]\\
&\quad +Q^{40}[({e_{10}}^2)/(1+\tau)+({m_{10}}^2\mu_0^2\tau)/((1+\tau))+2{m_{10}}^2\mu_0^2\tau\tan^2(\theta/2)]
\end{aligned}
$$

Proton Radius vs. Order of Polynomial Fits

MY Fits to the full 1422 points of the Mainz 2014 Data Using A Python Regression Code That Reproduces The PRC Results

- NOT THE SAME X^2 AS THE EARLIER FITS AS MAINZ FLOATS NORMALIZATIONS WITHOUT PENATLY TERMS -





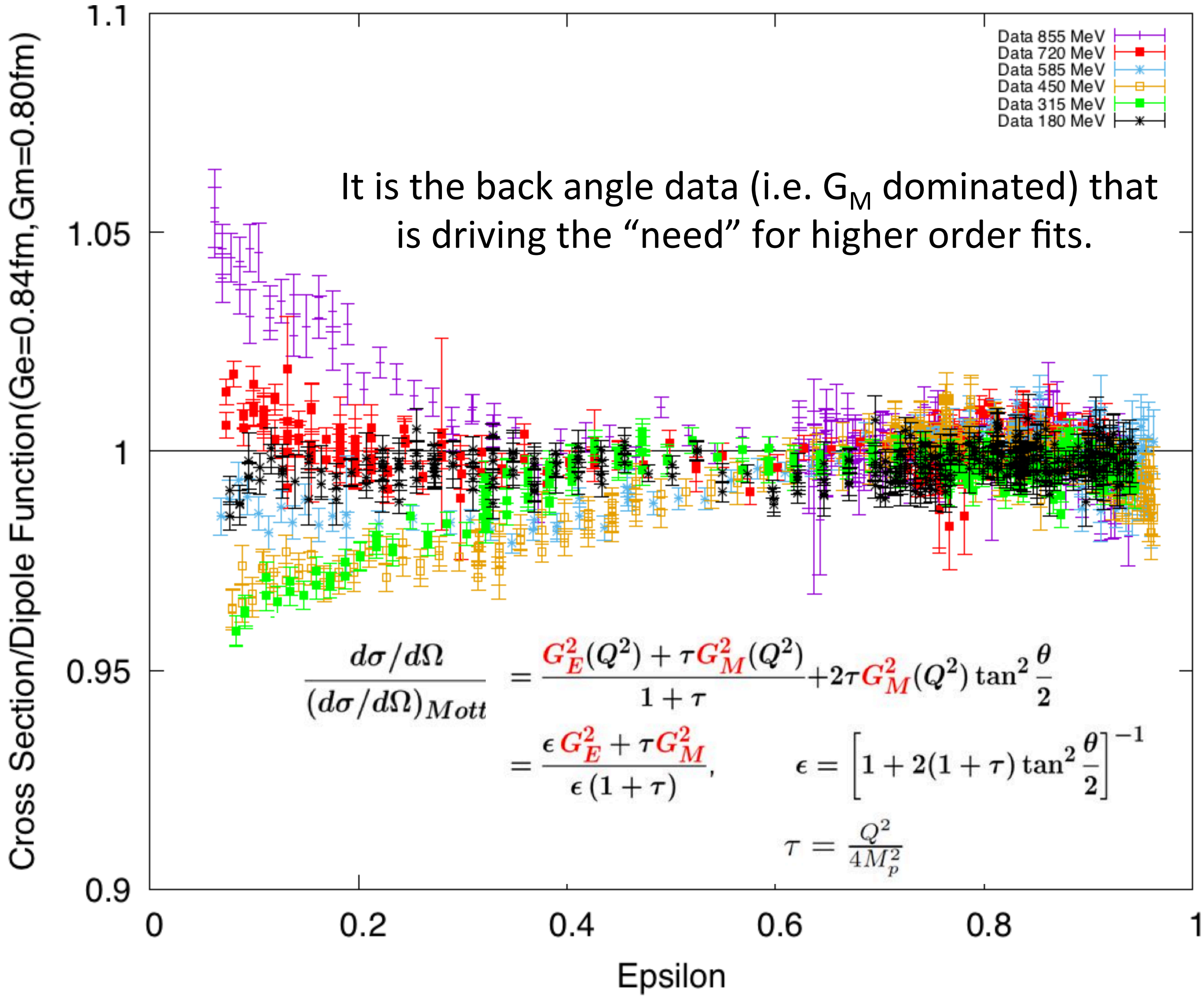
Fits all include the Mainz's 31 normalization parameters.

Even with 51 parameters (x2 10 & 31 norms) reduced chi2 never gets to unity.

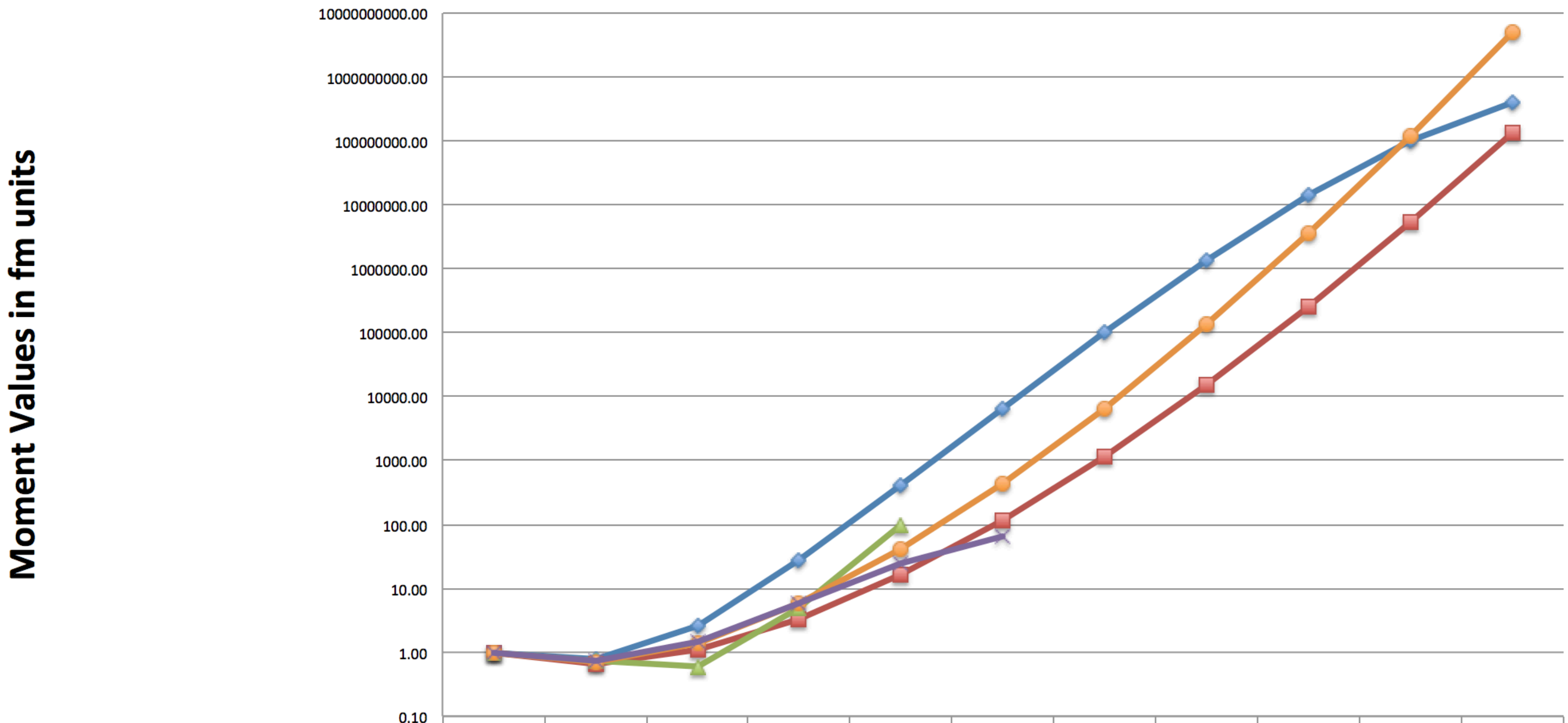
If uncertainty **estimates** are increased by 0.001, then 4th order reduced chi2 less then unity.

(see the Dos and don'ts of chi-square - http://arxiv.org/abs/1012.3754)

So what the heck is going on?! (residual to the dipole functions)



Disagreement About Moments Too

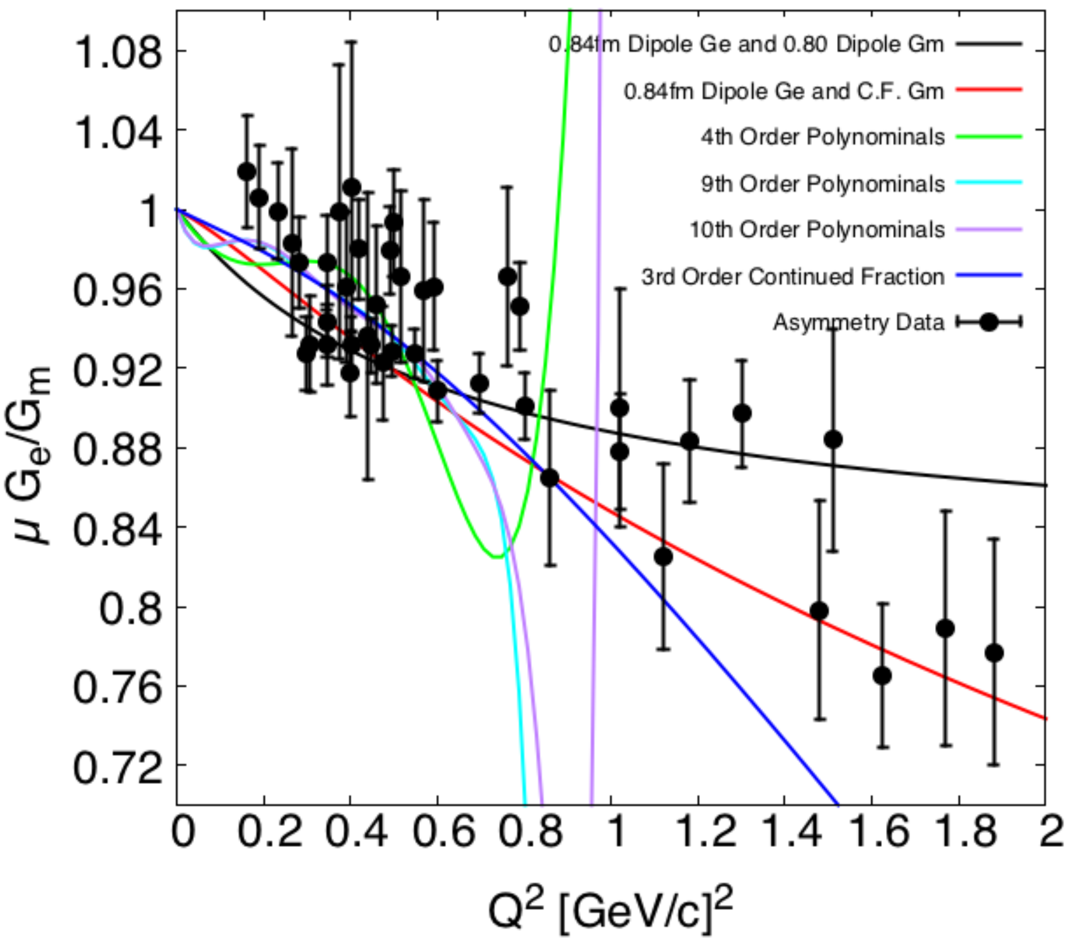


	(charge)	(radius)^2	<r^4>	<r^6>	<r^8>	<r^10>	<r^12>	<r^14>	<r^16>	<r^18>	<r^20>
Bernauer Moments	1.00	0.78	2.63	26.93	408.12	6597.29	1.00E+05	1.33E+06	1.40E+07	1.03E+08	3.99E+08
Standard Dipole Moments	1.00	0.66	1.08	3.32	16.37	118.40	1.18E+03	1.55E+04	2.60E+05	5.42E+06	1.37E+08
M. Horbatsch et al. PRC Power Series (N=4)	1.00	0.73	0.60	5.00	99.00						
Higinbotham et al., PRC Rational Function	1.00	0.70	1.38	5.68	40.06	431.90	6.60E+03	1.36E+05	3.62E+06	1.21E+08	5.00E+09
Higinbotham et al., PRC Power Series (N=5)	1.000	0.726	1.500	5.746	24.676	64.665					

The modern smaller radius fits agree with classic moments (e.g. I. Sick 2003)

Classic Test: Use data not included in the fit.

Fits of the full 1422 point Mainz using a Python fitting code based on the Mainz fitting routine.



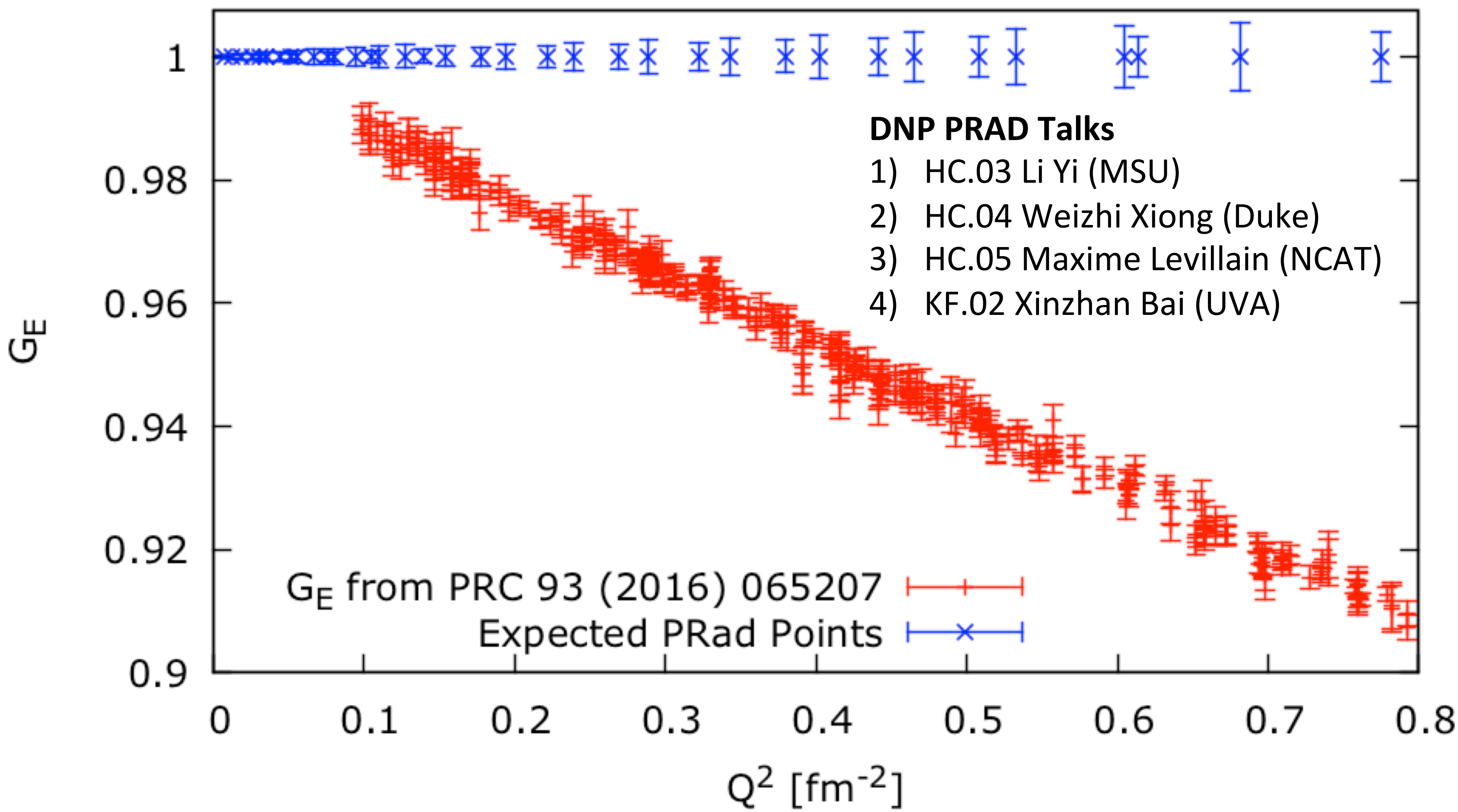
Function that can extrapolate well tend to give agreement with world G_M & a smaller radius…

PRad: Hall B Proton Radius Experiment

Small angle and small Q^2 to minimize the effects of G_M and provide best measurement of G_E
Gas Target (the proton) and GEM Detectors (scattering angles) & CEBAF (the beam energy)

The Collaboration & Jefferson Lab staff did an amazing job getting this experiment ready and was the first completed experiment for the upgraded CEBAF accelerator.

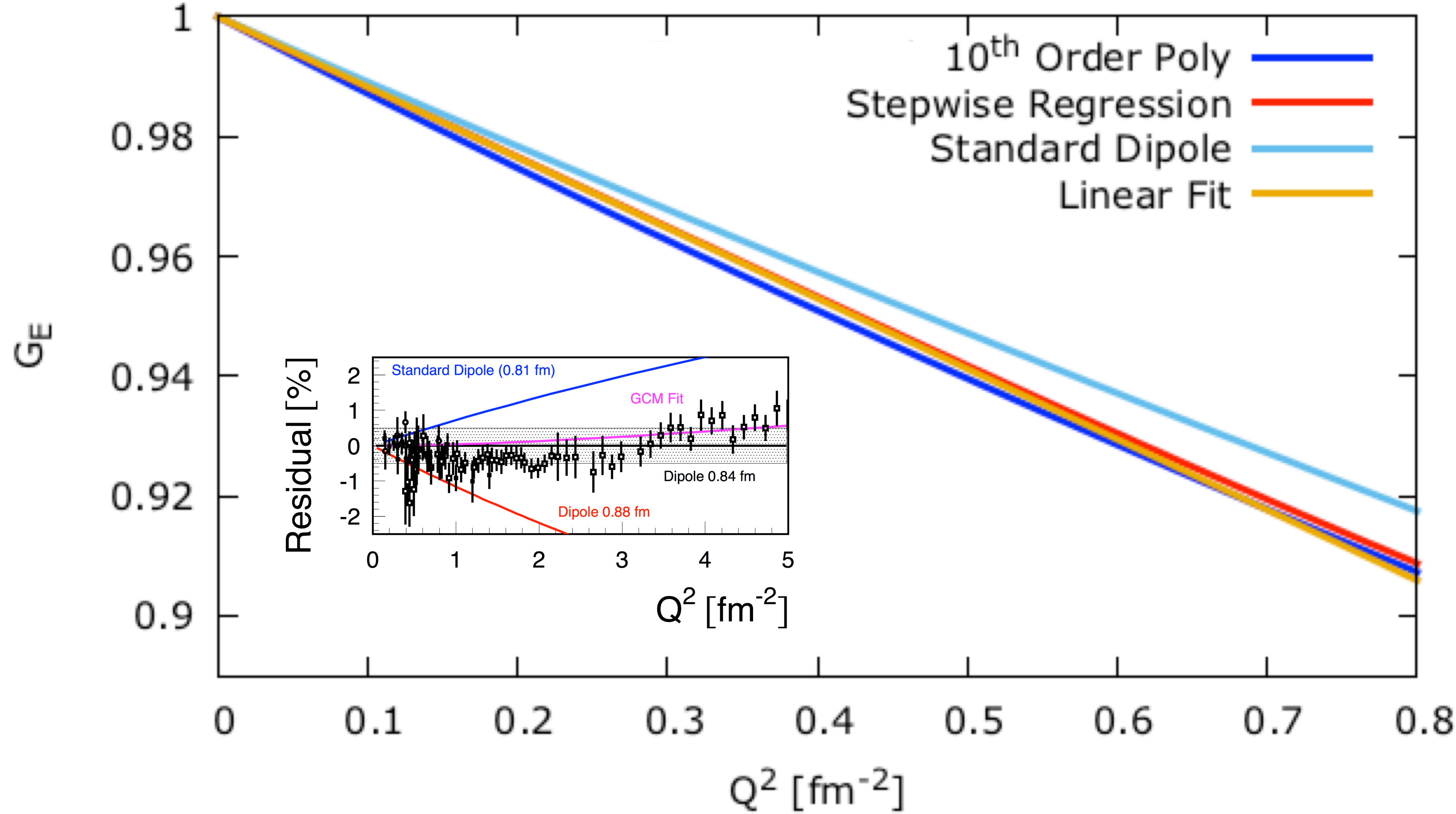
Expected Precision of PRad Data



DNP PRAD Talks

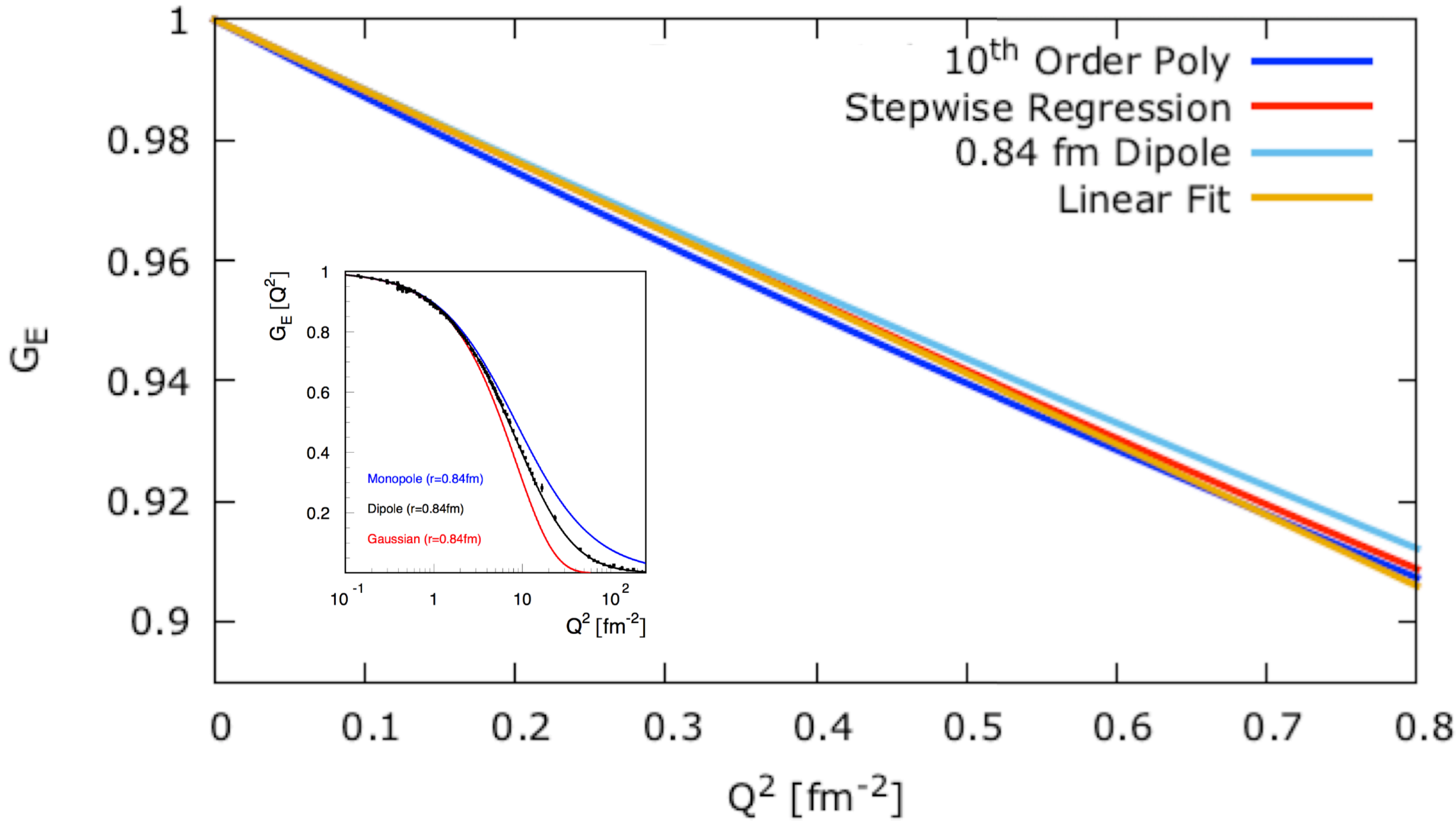
1) HC.03 Li Yi (MSU)
2) HC.04 Weizhi Xiong (Duke)
3) HC.05 Maxime Levillain (NCAT)
4) KF.02 Xinzhan Bai (UVA)

Simple G_E vs. Complex G_E



At least it is clear standard dipole doesn't work well . . .

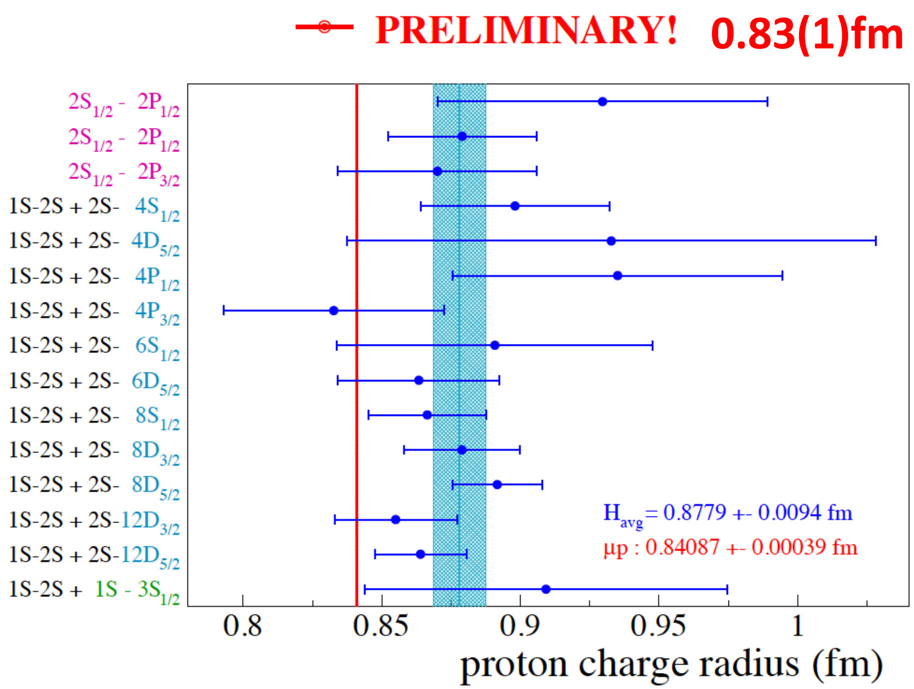
But Perhaps A Better "Standard" Dipole

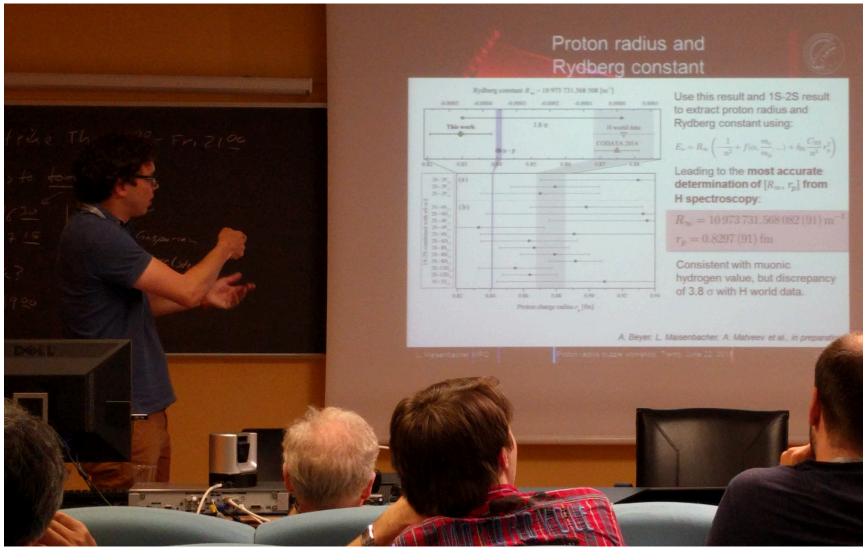


Summary

- **"All models are wrong, some models are"** – George Box
- **Occam's Razor** - *Among competing hypotheses, the one with the fewest assumptions should be selected.*
- **Confirmation Bias** - *Tendency to search for or interpret information in a way that confirms one's preconceptions.*
- To try to avoid confirmation bias, one can apply statistical modeling techniques such as F-tests, AIC, Stepwise Regression, etc. to determine the function to fit a given set of data.
  - R based Stepwise Regression Code Posted Along With Example Data Sets
  - http://jeffersonlab.github.io/model-selection/
- With these kinds of techniques, electron scattering data produces a proton radius consistent with the Muonic hydrogen data (0.84 fm)
  - Higinbotham et al., Phys. Rev. C **93** (2016) 055207
  - Griffioen, Carlson, and Maddox, Phys. Rev C **93** (2016) 065207 .
  - M. Horbatsch, E. A. Hessels, and A. Pineda, Phys. Rev. C 95 (2017) 035203.
- **Codes Such As R Provide An Amazing Open Source Statistical Toolbox!**
- Lots of new proton results coming, including PRad (Hall B), MUSE, ISR, as well as new Hydrogen Lamb shift measurements and perhaps a new Rydberg constant…

New Atomic Hydrogen Lamb Data





Result has been submitted for publication.

Preliminary results from talk given at HC2NP and photos from Trento workshop.
for details see Pohl's talk at https://indico.cern.ch/event/492464/timetable/#20160930.detailed

Further Reading

- Particle Data Handbook – Statistics Section
  - *http://pdg.lbl.gov/2015/reviews/rpp2015-rev-statistics.pdf*
- The Interpretation of Errors – Fredrick James
  - *http://seal.cern.ch/documents/**minuit**/mn**error**.pdf*
- Data Analysis Textbooks
  - Data Reduction and Error Analysis – Philip Bevington
  - Statistical Methods in Experimental Physics – Fredrick James
  - Computation Methods for the Physical Science – Simon Širca
  - Probability of Physics – Simon Širca
- R Programing Language
  - https://www.r-project.org/
- **Estimation**
  - Street-Fighting Mathematics – Sanjoy Mahajan
  - Guesstimation – Larry Weinstein