Our treatment of systematic errors

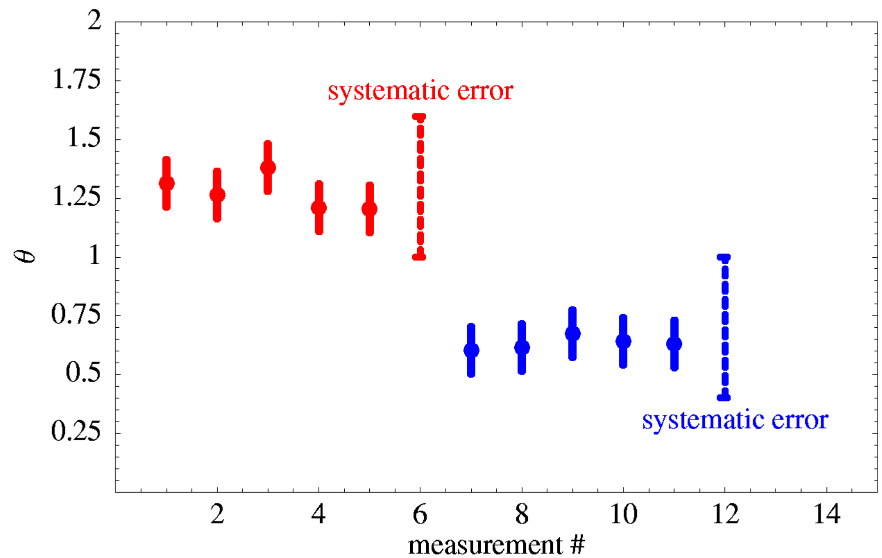
What is a systematic error?

"This is why people are so frightened of systematic errors, and most other textbooks avoid the subject altogether. You never know whether you have got them and can never be sure that you have not – like an insidious disease…

The good news, however, is that despite popular prejudices and superstitions, once you know what your systematic errors are, they can be handled with standard statistical methods."

R. J. Barlow
Statistics

Imagine that two experimental groups have measured a quantity θ, with the results shown.



OK, what is the value of θ?

This is very analogous to what happens in global analysis of PDF's. But in the case of PDF's the systematic differences are only visible through the PDF's.

We use χ^2 minimization with fitting of systematic errors.

For *statistical errors* define

$$\chi^2 = \sum_{i=1}^{N} \frac{(D_i - T_i)^2}{\sigma_i^2} \qquad \begin{cases} D_i : \text{data value} \\ T_i : \text{theoretical value} \\ \sigma_i : \text{statistical error} \end{cases}$$

(S. D.)

$T_i = T_i(a_1, a_2, \ldots, a_d)$ a function of d theory parameters

Minimize χ^2 w. r. t. $\{a_\mu\}$ $\Rightarrow$ optimal parameter values $\{a_{0\mu}\}$.

All this would be based on the assumption that

$$D_i = T_i(a_0) + \sigma_i r_i$$

$$dP = \frac{e^{-r^2/2}}{\sqrt{2\pi}} dr$$

Treatment of the normalization error

In scattering experiments there is an overall normalization uncertainty from uncertainty of the luminosity. We define

$$\chi^2(a, f_{\mathrm{N}}) = \left(\frac{1 - f_{\mathrm{N}}}{\sigma_{\mathrm{norm}}}\right)^2 + \sum_{i=1}^{N} \frac{(f_{\mathrm{N}} D_i - T_i)^2}{\sigma_i^2}$$

where f_{N} = overall normalization factor

Minimize χ^2 w. r. t. both $\{a_\mu\}$ and f_{N}.

A method for general systematic errors

$$D_i = T_i(a_0) + \alpha_i r_i + \sum_{j=1}^{K} \beta_{ij} \hat{r}_j$$

α_i : statistical error of D_i
β_{ij} : set of systematic errors ($j=1\ldots K$) of D_i

Define

$$\chi'^2 = \sum_{i=1}^{N} \frac{\left(D_i - \sum_j \beta_{ij} s_j - T_i\right)^2}{\alpha_i^2} + \sum_{j=1}^{K} s_j^2$$

quadratic penalty term

Minimize χ'^2 with respect to both shape parameters $\{a_\mu\}$ and optimized systematic shifts $\{s_j\}$.

Because $\chi'2$ depends quadratically on $\{s_j\}$ we can solve for the systematic shifts analytically

$$\frac{\partial \chi'^2}{\partial s_j} = 0 \quad \text{implies} \quad s_j = s_j^{(0)}(a) = \sum_{m=1}^{K} \left(A^{-1}\right)_{jm} B_m$$

where

$$A_{jk} = \delta_{jk} + \sum_{i=1}^{N} \frac{\beta_{ij}\beta_{ik}}{\alpha_i^2} \quad \text{and} \quad B_j = \sum_{i=1}^{N} \frac{(D_i - T_i)\beta_{ij}}{\alpha_i^2}$$

K x *K* matrix   *K* x *1* matrix

Now let

$$\chi^2_{\text{global}}(a) = \sum_{\text{experiments}} \chi'^2\left(a, s_0(a)\right)$$

and minimize w.r.t $\{a_\mu\}$.

The systematic shifts $\{s_j\}$ are continually optimized
[$\mathbf{s} \rightarrow \mathbf{s}_0(a)$]

So, we have accounted for …

- Statistical errors
- Overall normalization uncertainty (by fitting $\{f_{\mathrm{N},e}\}$)
- Other systematic errors (analytically)

We may make further refinements of the fit with weighting factors

$$\chi^2_{\mathrm{global}}(\{a\},\{f_{\mathrm{N}}\}) = \sum_e w_e \chi^2_e(\{a\},\{f_{\mathrm{N}}\}) + \sum_e w_{\mathrm{N},e}\left(\frac{(1-f_e)^2}{\sigma^2_{\mathrm{N},e}}\right)$$

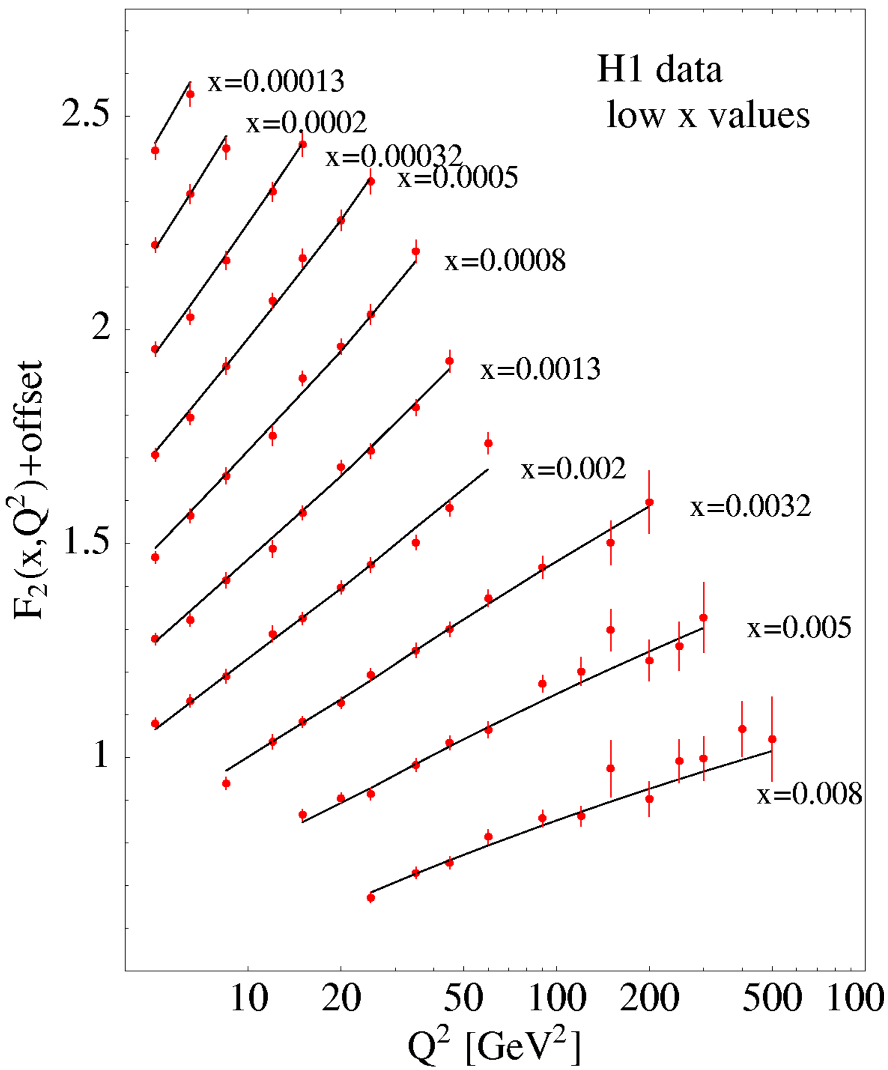
Default : w_e and $w_{\mathrm{N},e} = 1$

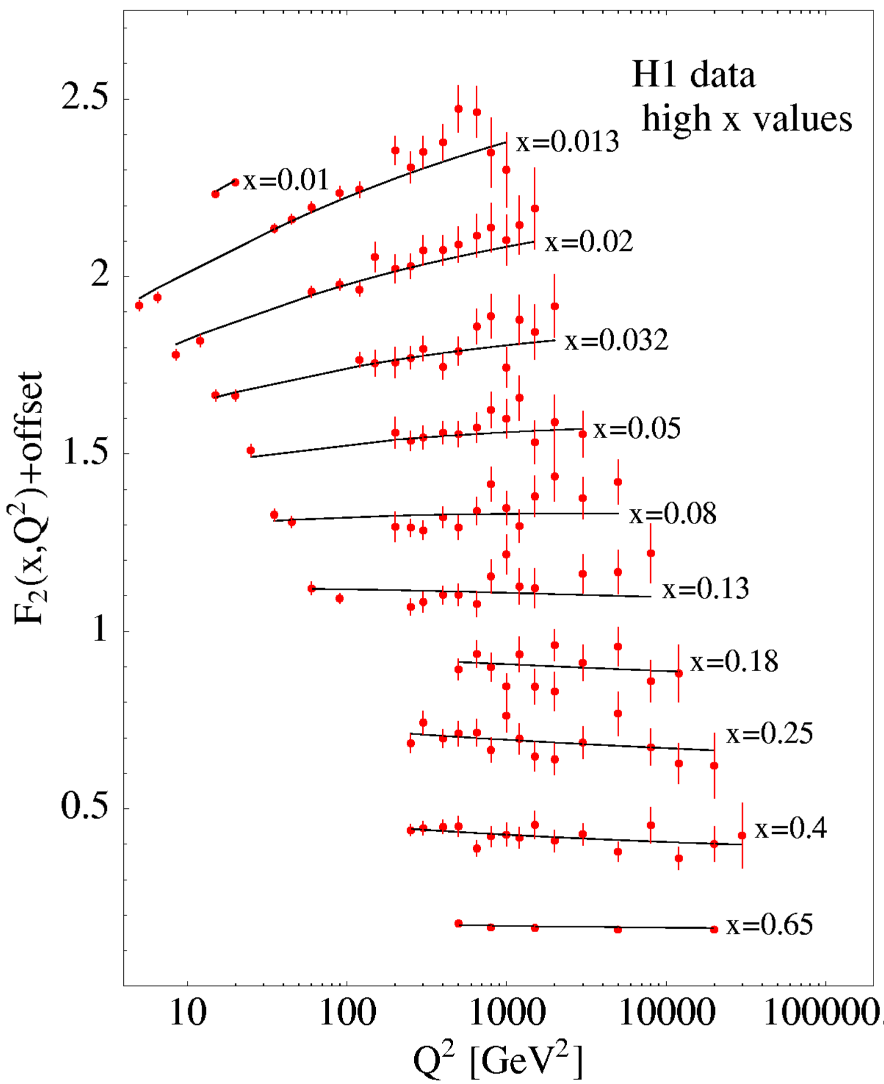
The spirit of global analysis is compromise – the PDF's should fit all data sets satisfactorily.
If the default leaves some experiments unsatisfied, we may be willing to reduce the quality of fit to some experiments in order to fit better another experiment. (We use this sparingly!)

Quality

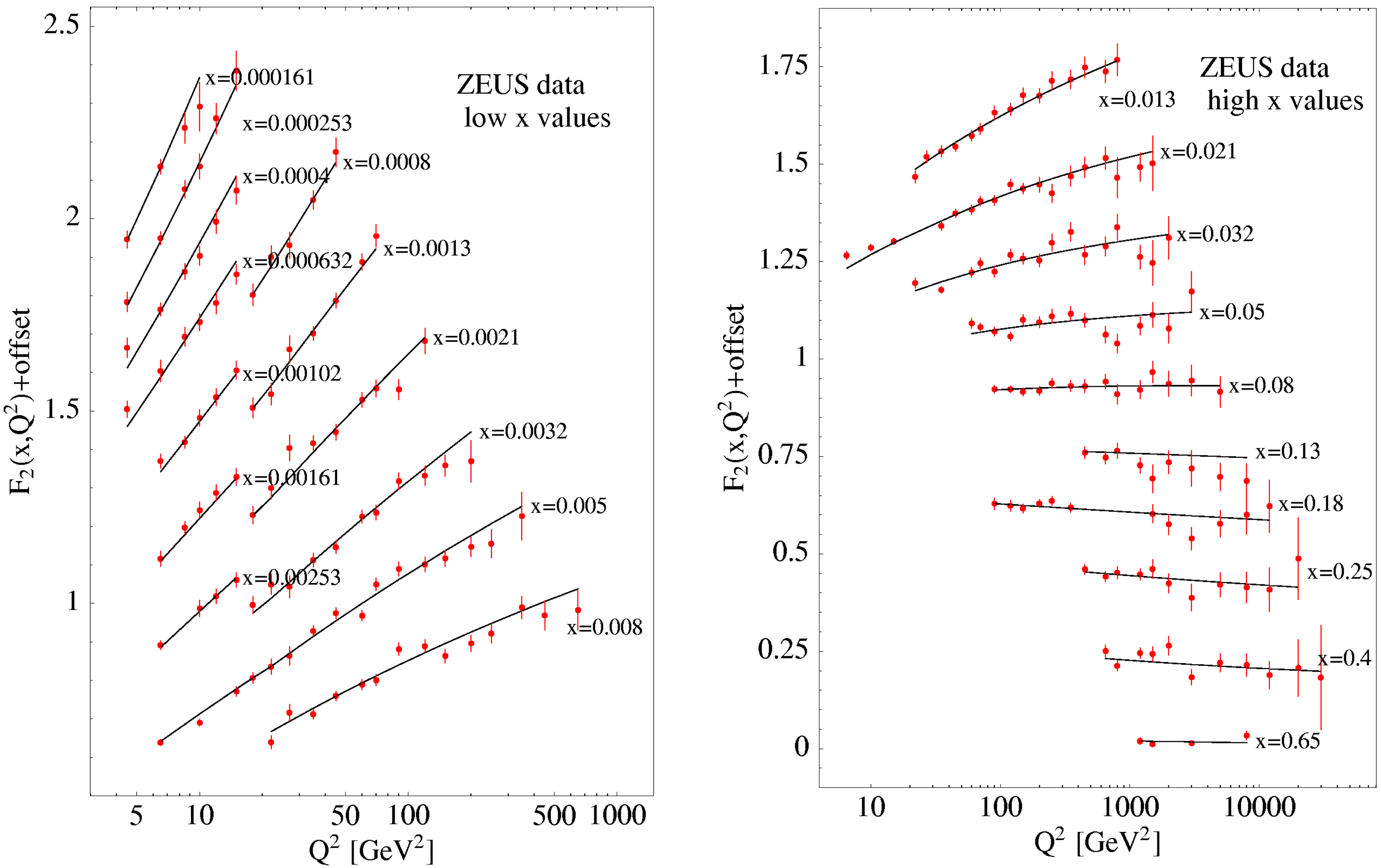
How well does this fitting procedure work?

Comparison of the CTEQ6M fit to the H1 data in separate x bins. The data points include optimized shifts for systematic errors. The error bars are statistical only.

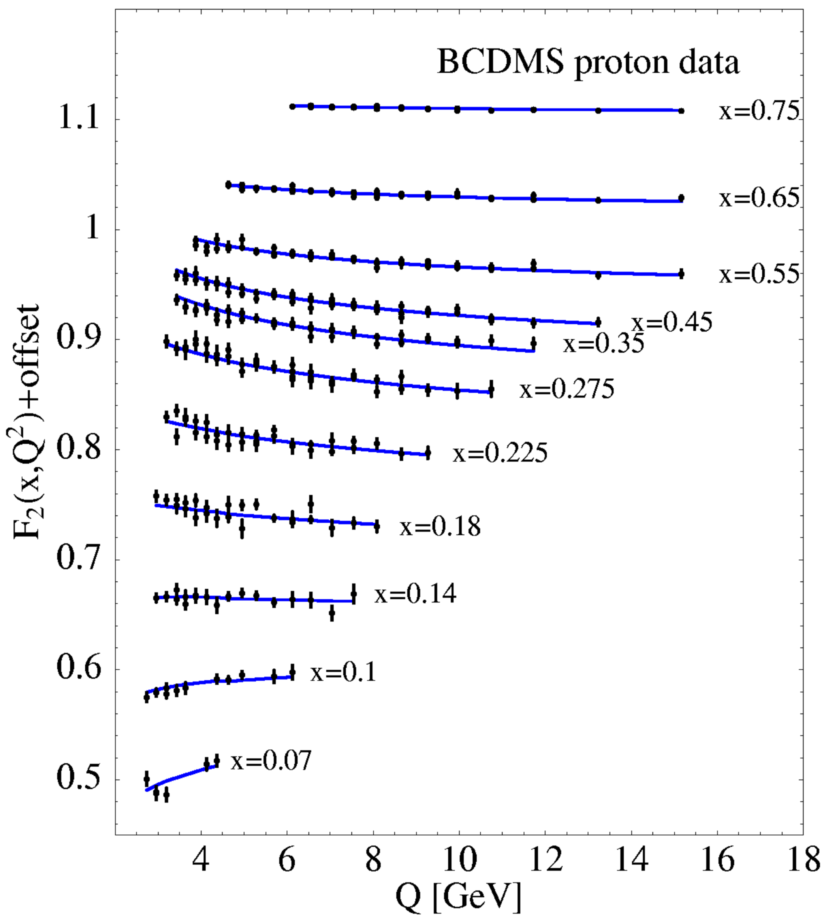


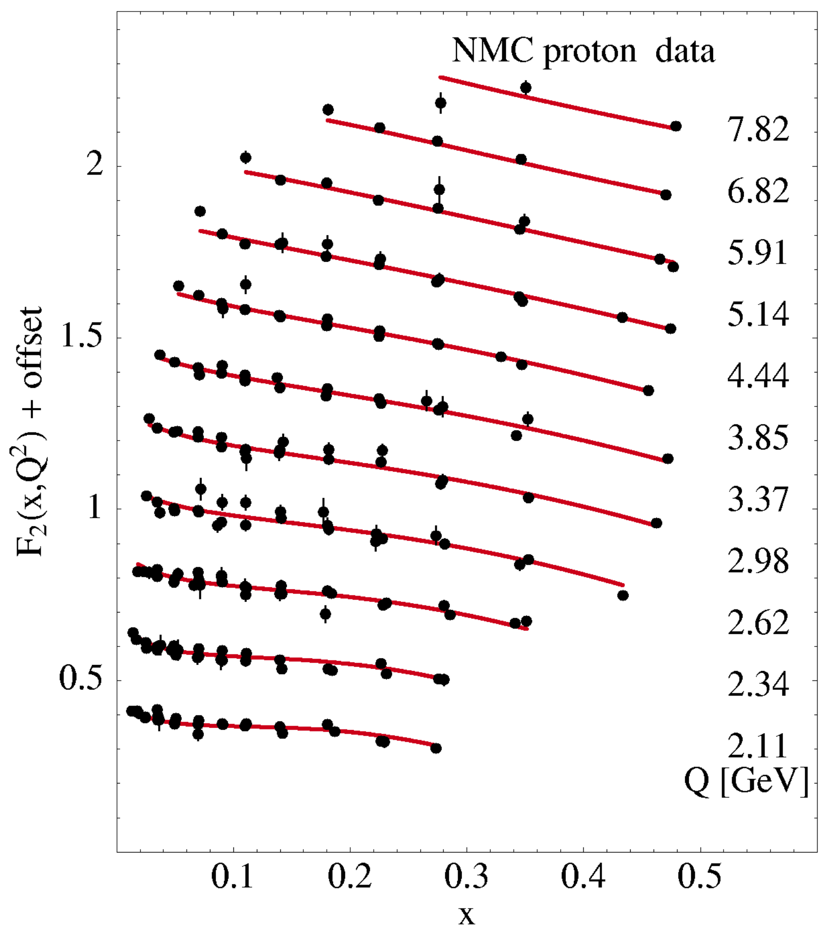


Comparison of the CTEQ6M fit to the ZEUS data in separate x bins. The data points include optimized shifts for systematic errors. The error bars are statistical only.

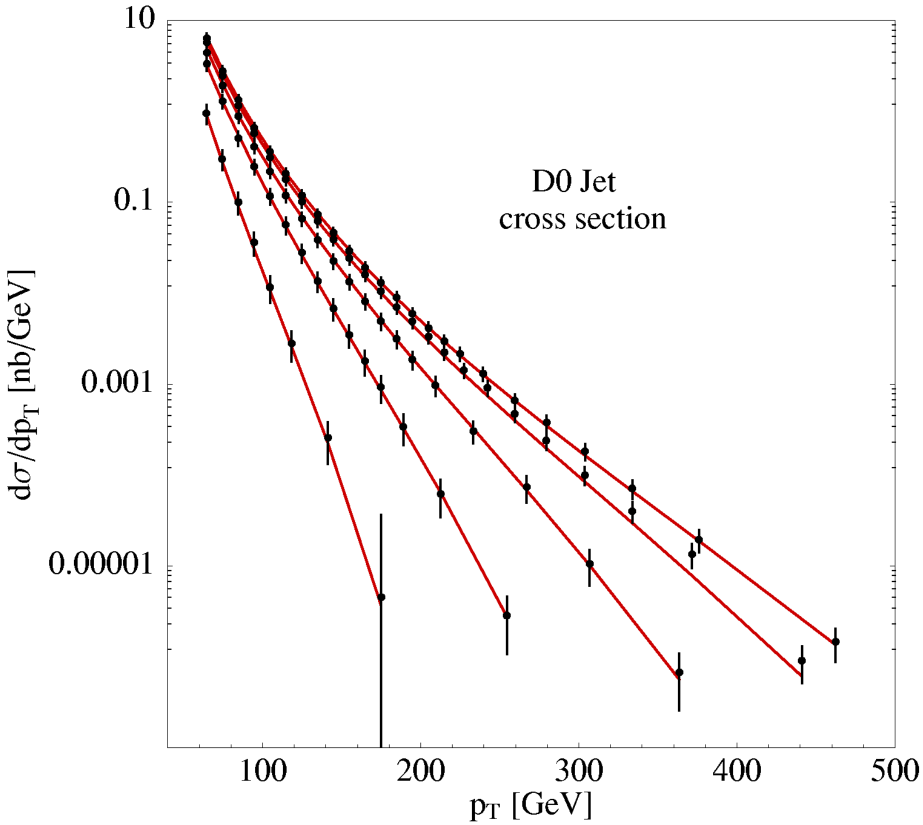


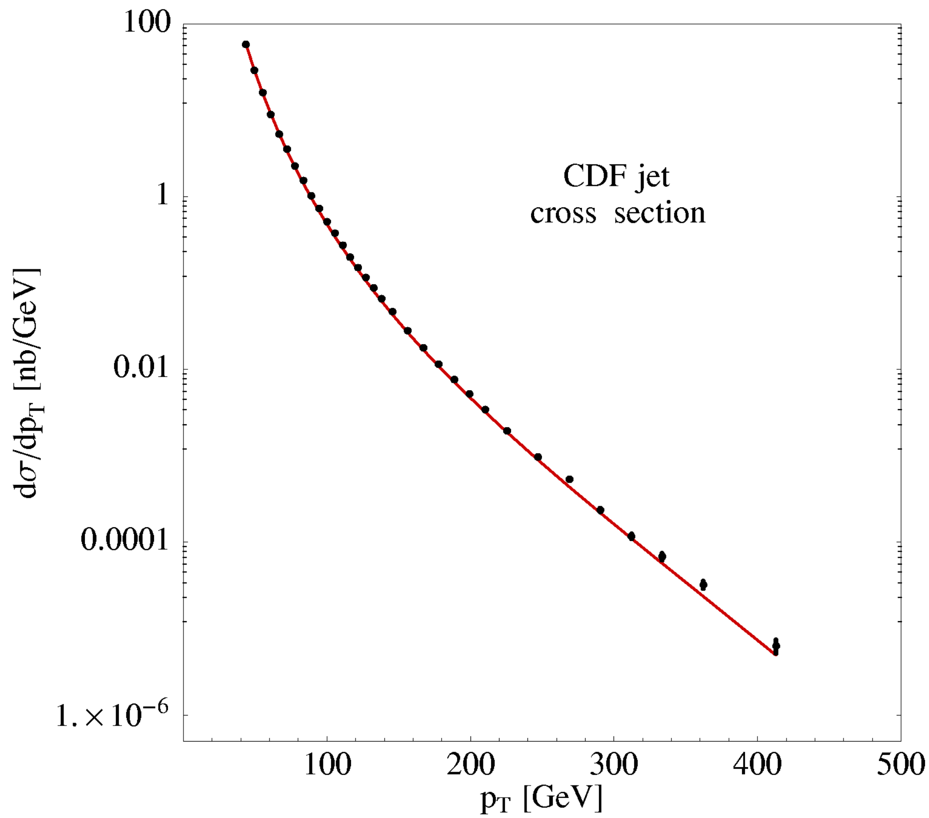
Comparison of the CTEQ6M fit to the BCDMS and NMC data on μp scattering. The data points include optimized shifts for systematic errors. The error bars are statistical only.



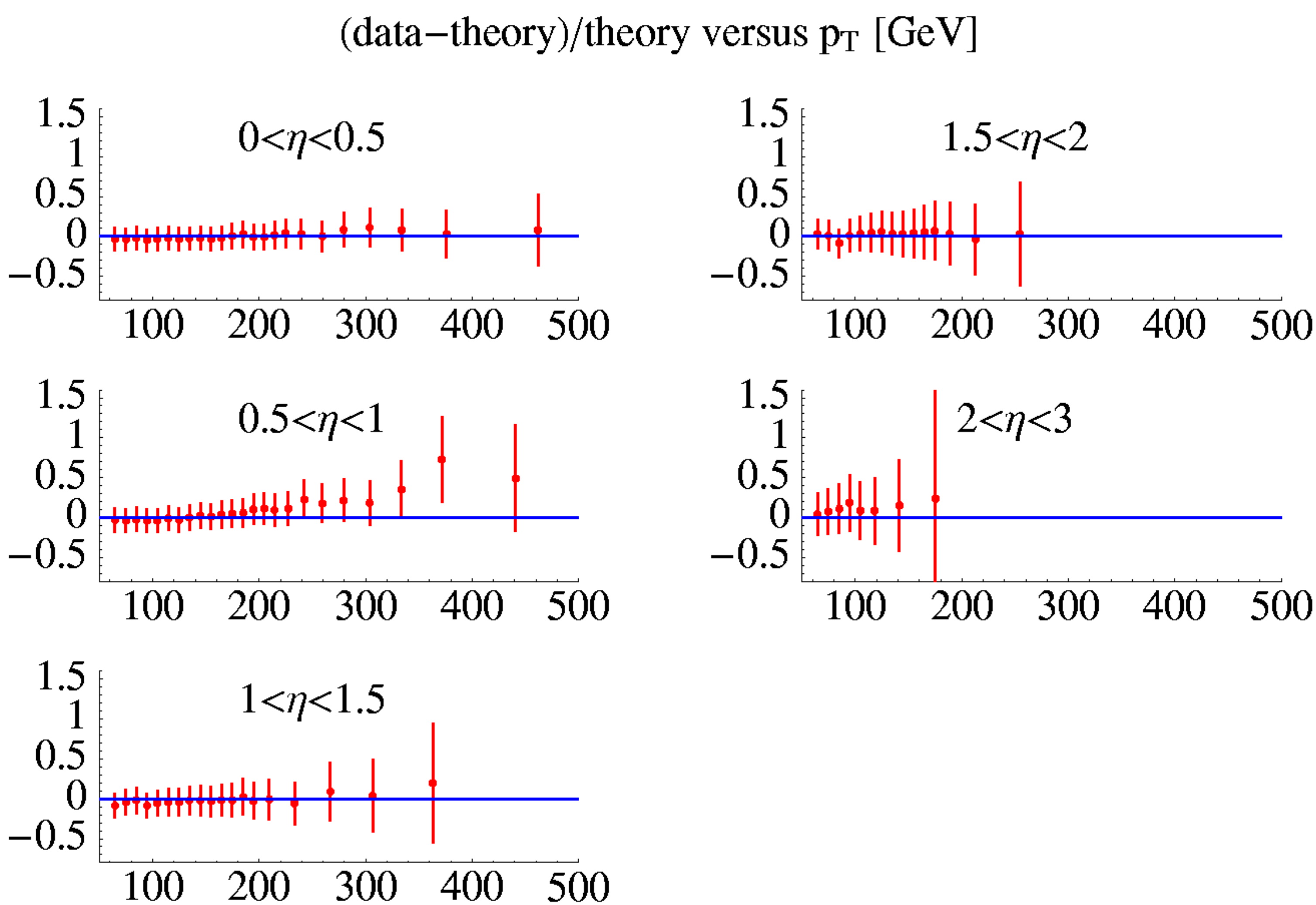


Comparison of the CTEQ6M fit to the inclusive jet data. (a) D0 cross section versus p_T for 5 rapidity bins; (b) CDF cross section for central rapidity.





Closer comparison between CTEQ6M and the D0 jet data as fractional differences.



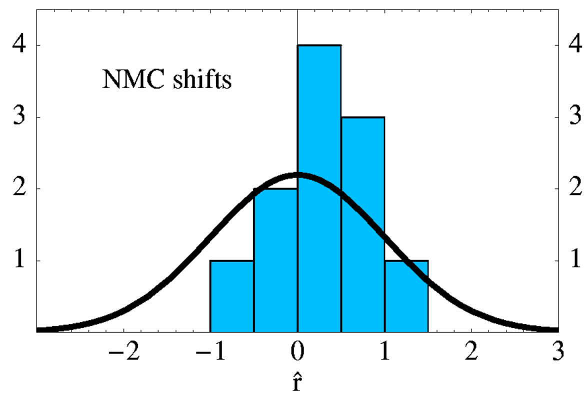
How large are the optimized normalization factors?

Expt	f_N
BCDMS	0.976
H1 (a)	1.010
H1 (b)	0.988
ZEUS	0.997
NMC	1.011
CCFR	1.020
E605	0.950
D0	0.974
CDF	1.004

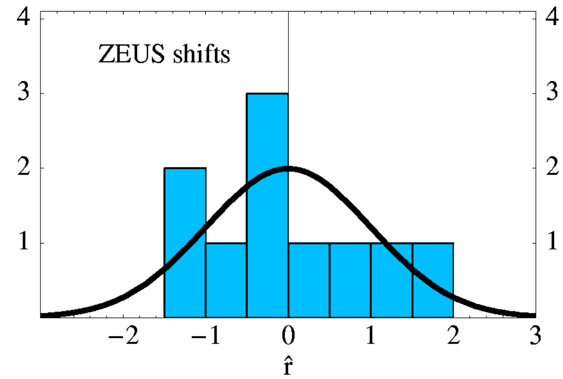
We must always check that the systematic shifts are not unreasonably large.

10 systematic shifts NMC data

j	sj
1	1.67
2	-0.67
3	-1.25
4	-0.44
5	0.00
6	-1.07
7	1.28
8	0.62
9	-0.40
10	0.21

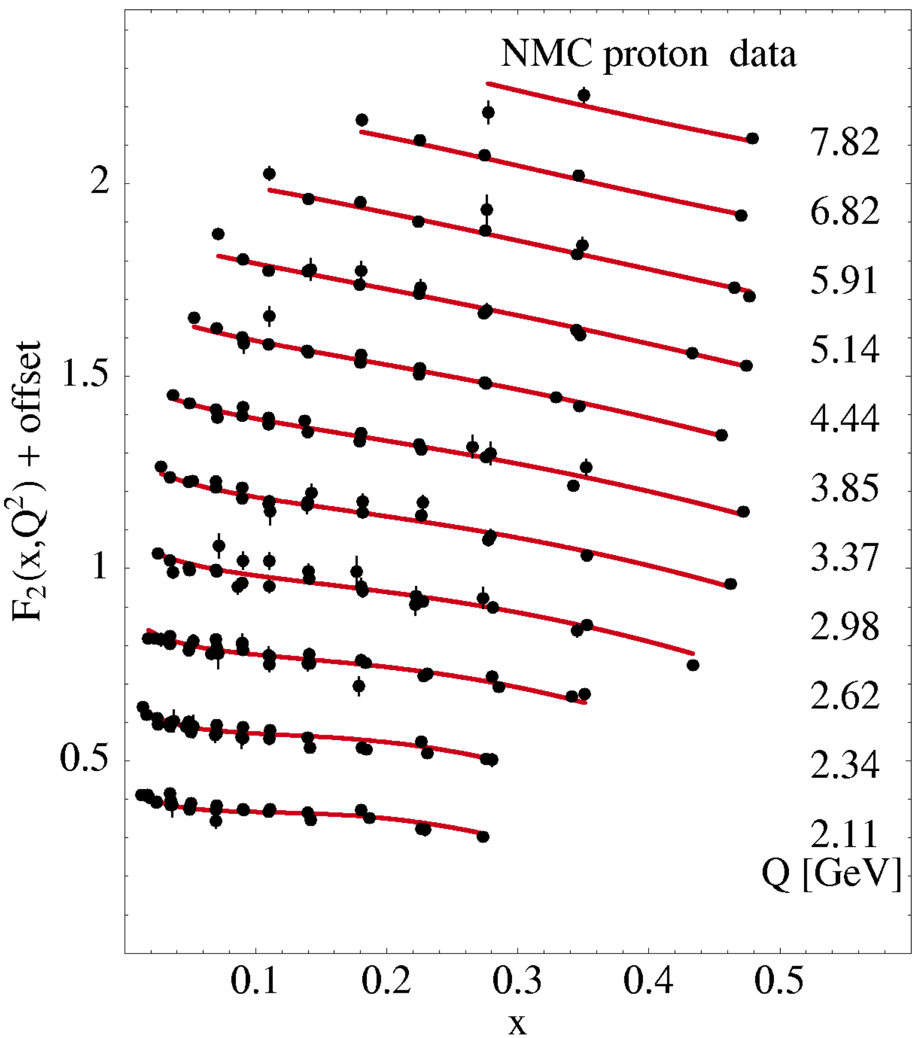


11 systematic shifts ZEUS data

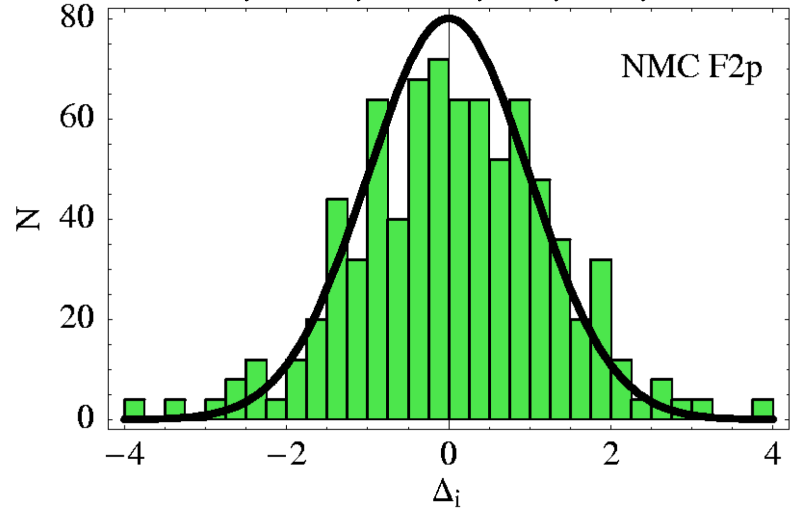


j	sj
1	0.67
2	-0.81
3	-0.35
4	0.25
5	0.05
6	0.70
7	-0.31
8	1.05
9	0.61
10	0.26
11	0.22

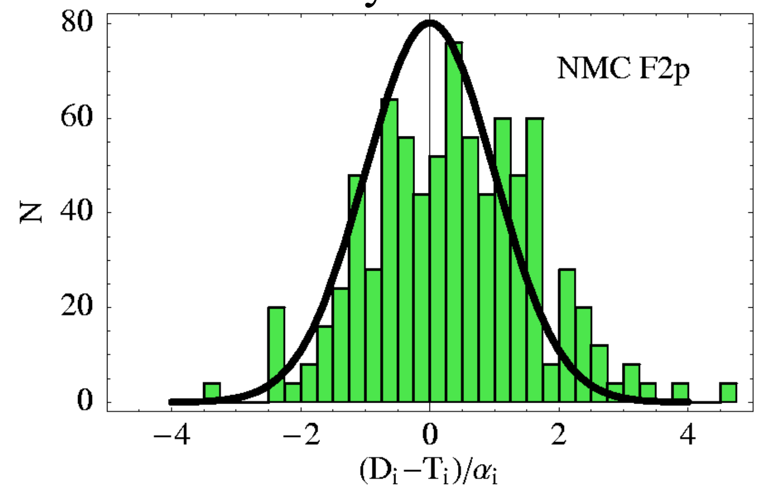
Comparison to NMC F_2



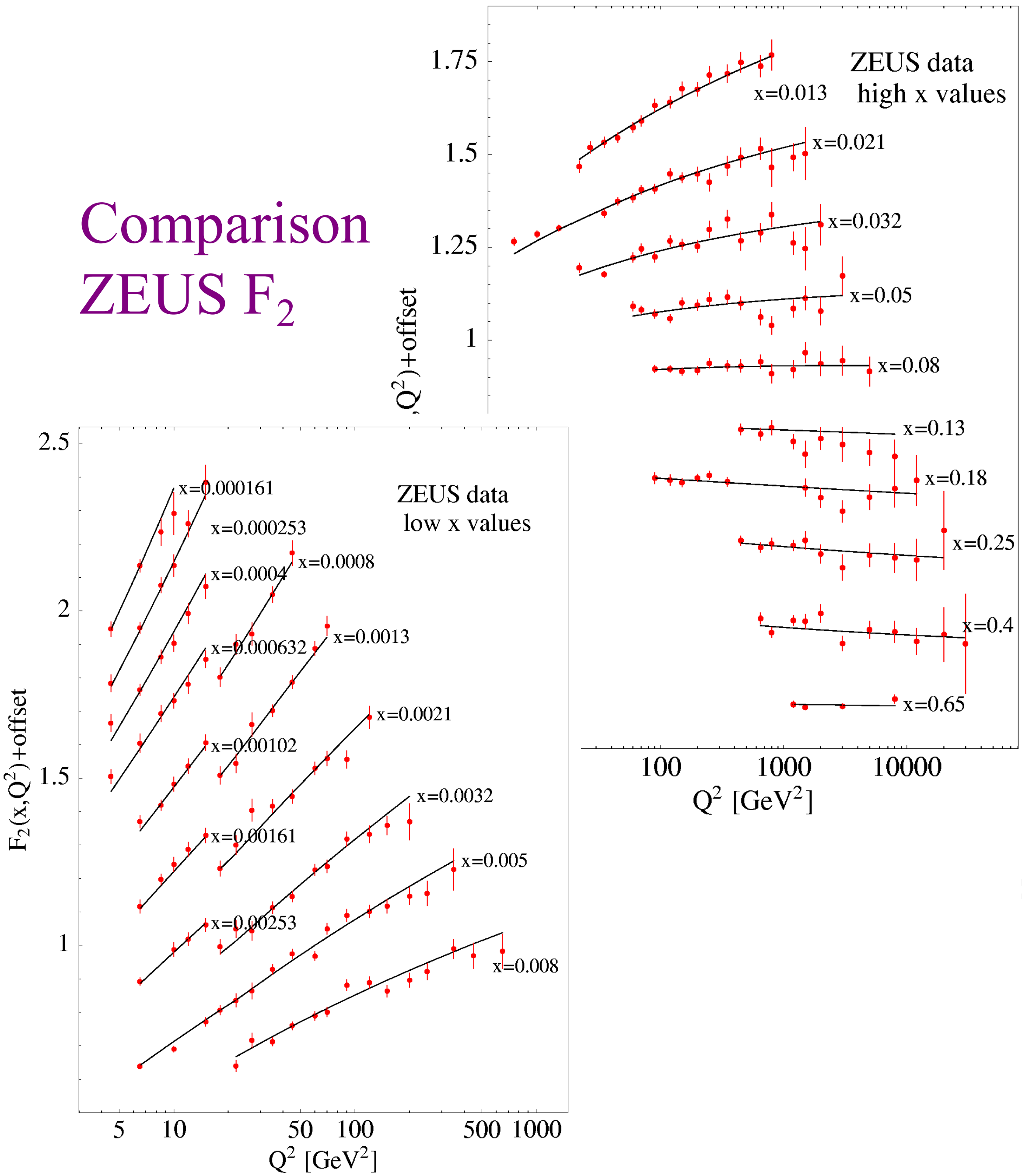
$$\Delta_i = (D_i - \mathbf{s}\cdot\boldsymbol{\beta}_i - T_i)/\alpha_i$$



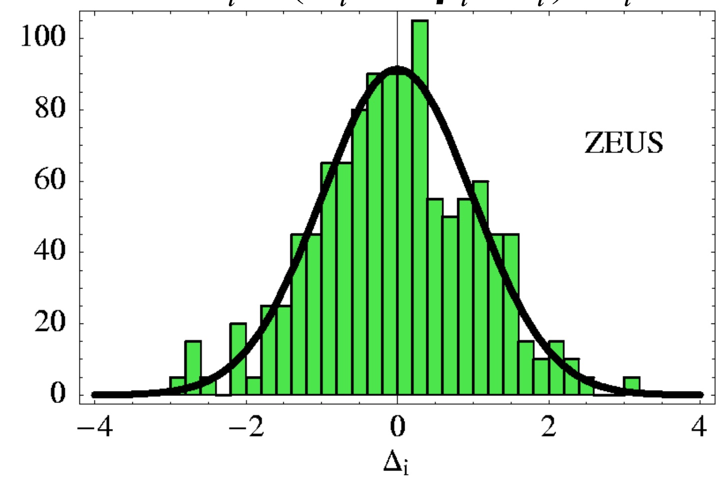
without systematic shifts



Comparison ZEUS F_2



$$\Delta_i = (D_i - \mathbf{s}\cdot\boldsymbol{\beta}_i - T_i)/\alpha_i$$



without systematic shifts